Biochemical Studies of Postsynaptic Density Signaling Proteins With a Focus on synGAP and PDZ Domains

Thesis by
Ward Gale Walkup IV

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California
2014
(Defended December 16th, 2013)

ACKNOWLEDGEMENTS

First and foremost I would like to thank my mother and "Second Ph.D. Advisor", Dr. Kathryn Ivanetich, for introducing me to science and for helping me through each step of my career, from counting spools of thread to learning arithmetic in the first grade to discussing the catalytic mechanisms of polyketide synthases. Mom, I couldn't have made it this far without you, and am forever grateful for your help, compassion, ability to crack the whip, and willingness to take your son surfing at the drop of a hat.

I am extremely grateful to Mary B. Kennedy for her mentoring and support. As you've told me many times, "Good science is hard and takes time," and I would like to extend my utmost gratitude for your scientific and financial support throughout the peaks and valleys of my thesis research. I appreciate all of the brutally honest and constructive feedback you've provided, and I am proud to have worked for such an excellent scientist. I would also like to thank you for providing the creative freedom and lab reagents to investigate topics I found intellectually stimulating, and for reining me in when I went off track or delved too far into minutia.

I would never have successfully completed my thesis research without the help and prolific insight of Dr.'s Mark Welch (DNA2.0) and Jost Vielmetter (Caltech Protein Expression Center). Mark, thank you for being willing to discuss (*ad nauseam*) the important things in life, whether it was a recent swing or equipment change, something stupid I did (e.g. that time where I accidentally ingested PKS lysate as it shot out of the French Press), or the optimal coding sequence for a heterologously expressed neuronal gene. Jost, thank you teaching me the ins and outs of protein expression and purification,

including the proper way to pack and run a size exclusion column, and for always being there to fix a piece of equipment when something in the lab broke.

I would like to thank the following former and current Mary B. Kennedy lab members for their support, both technical and emotional: Leslie Schenker, Melanie Stefan, Holly Carlisle, Eduardo "Dado" Marcora, Tami Kinzer-Ursem, and Barry "Dylan" Bannon.

I would like to thank the National Science Foundation (Graduate Fellowship & Interdisciplinary Post Doctoral Fellowship) and National Institutes of Health (NIH) (NIH Trainee) for their past, present, and future financial support. I would also like to thank Liz Ayala and Paul Sternberg for approving all of my supplemental supply allowance requests, especially when funding was tight.

I am also grateful to the members of my thesis committee, including Douglas C. Rees, Steven L. Mayo, Pamela J. Bjorkman, and Mary B. Kennedy, for their scientific advice and patience during my incredibly long and arduous road to fulfill the requirements for a Caltech Ph.D.

I would like to thank Michael Anaya, Geoffrey Lovely, Erik Schomburg, Alon Gorodetsky, Phil and Alex Romero, and Russ Ernst for making my time at Caltech that much more entertaining, and Len "Lanny" Bourcy, Dr. David "I Have No Edit Button" Barba, Matt "Scumbag" Woodworth, and Andrew "Meathead" McDade for all the memorable moments and sheer foulness over the years.

I would like to thank Queen Elizabeth Cody-Walkup and Prince Charming Cody-Walkup for providing comic relief, kisses, and cuddling when I was feeling less than 100%. You two are the best fur children a parent could ask for.

Finally, I would like to thank my wife, Julia Cody Walkup, for everything, as I can't imagine my life without you… thank you for the emotional support, cheeseburgers, being the first woman to call me on my transgressions, finding my numerous personality quirks and foul mouth endearing, planning and taking care of all the things I let slip through the cracks, and for all of the fun and ridiculous memories we've shared. And last but not least, thank you for helping to produce an insanely cute and gaseous baby girl.

ABSTRACT

Memory storage in the brain involves adjustment of the strength of existing synapses and formation of new neural networks. A key process underlying memory formation is synaptic plasticity, the ability of excitatory synapses to strengthen or weaken their connections in response to patterns of activity between their connected neurons. Synaptic plasticity is governed by the precise pattern of Ca^{2+} influx through postsynaptic N-methyl-D-aspartate-type glutamate receptors (NMDARs), which can lead to the activation of the small GTPases Ras and Rap. Differential activation of Ras and Rap acts to modulate synaptic strength by promoting the insertion or removal of 2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid receptors (AMPARs) from the synapse. Synaptic GTPase activating protein (synGAP) regulates AMPAR levels by catalyzing the inactivation of GTP-bound (active) Ras or Rap. synGAP is positioned in close proximity to the cytoplasmic tail regions of the NMDAR through its association with the PDZ domains of PSD-95. SynGAP's activity is regulated by the prominent postsynaptic protein kinase, Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) and cyclin-dependent kinase 5 (CDK5), a known binding partner of CaMKII. Modulation of synGAP's activity by phosphorylation may alter the ratio of active Ras to Rap in spines, thus pushing the spine towards the insertion or removal of AMPARs, subsequently strengthening or weakening the synapse. To date, all biochemical studies of the regulation of synGAP activity by protein kinases have utilized impure preparations of membrane bound synGAP. Here we have clarified the effects of phosphorylation of synGAP on its Ras and Rap GAP activities by preparing and utilizing purified, soluble recombinant synGAP, Ras, Rap, CaMKII, CDK5, PLK2, and CaM. Using mass spectrometry, we have confirmed the

presence of previously identified CaMKII and CDK5 sites in synGAP, and have identified novel sites of phosphorylation by CaMKII, CDK5, and PLK2. We have shown that the net effect of phosphorylation of synGAP by CaMKII, CDK5, and PLK2 is an increase in its GAP activity toward HRas and Rap1. In contrast, there is no effect on its GAP activity toward Rap2. Additionally, by assaying the GAP activity of phosphomimetic synGAP mutants, we have been able to hypothesize the effects of CDK5 phosphorylation at specific sites in synGAP. In the course of this work, we also found, unexpectedly, that synGAP is itself a Ca^{2+}/CaM binding protein. While Ca^{2+}/CaM binding does not directly affect synGAP activity, it causes a conformational change in synGAP that increases the rate of its phosphorylation and exposes additional phosphorylation sites that are inaccessible in the absence of Ca2+/CaM.

The postsynaptic density (PSD) is an electron-dense region in excitatory postsynaptic neurons that contains a high concentration of glutamate receptors, cytoskeletal proteins, and associated signaling enzymes. Within the PSD, three major classes of scaffolding molecules function to organize signaling enzymes and glutamate receptors. PDZ domains present in the Shank and PSD-95 scaffolds families serve to physically link AMPARs and NMDARs to signaling molecules in the PSD. Because of the specificity and high affinity of PDZ domains for their ligands, I reasoned that these interacting pairs could provide the core components of an affinity chromatography system, including affinity resins, affinity tags, and elution agents. I show that affinity columns containing the PDZ domains of PSD-95 can be used to purify active PDZ domain-binding proteins to very high purity in a single step. Five heterologously expressed neuronal proteins containing

endogenous PDZ domain ligands (NMDAR GluN2B subunit Tail, synGAP, neuronal nitric oxide synthase PDZ domain, cysteine rich interactor of PDZ three and cypin) were purified using PDZ domain resin, with synthetic peptides having the sequences of cognate PDZ domain ligands used as elution agents. I also show that conjugation of PDZ domain-related affinity tags to Proteins Of Interest (POIs) that do not contain endogenous PDZ domains or ligands does not alter protein activity and enables purification of the POIs on PDZ domain-related affinity resins.

TABLE OF CONTENTS

Chapter IV:
Substrate Directed Regulation of synGAP Phosphorylation by Calcium/Calmodulin

Appendix I:
Mass Spectra for CamKII, CDK5 and PLK2 Phosphorylation Sites in synGAP

Appendix II:
Publications

LIST OF FIGURES AND TABLES

Chapter I:
Overview of Synaptic Signaling

Chapter II:
PDZ Affinity Chromatography: A Method for Affinity Purification of Proteins Based on PDZ Domains and Their Ligands

Chapter IV:
Substrate Directed Regulation of synGAP Phosphorylation by Calcium/Calmodulin

Chapter I

OVERVIEW OF SYNAPTIC SIGNALING

Spines, Synapses, and Synaptic Plasticity

Memory storage in the brain involves adjustment of the strength of existing synapses and formation of new neural networks. A key process underlying memory formation is synaptic plasticity, which is the ability of excitatory synapses to strengthen or weaken their connections in response to patterns of activity between their connected neurons [1]. In the brain, most synapses onto excitatory neurons are located on spines, small membranous protrusions from a neuron's highly branched dendrites. A typical excitatory neuron in the hippocampus makes approximately 10,000 connections via its spines, with most spines hosting a single synapse. The size of postsynaptic spines varies in shape from thin to stubby or mushroom shaped, with stronger synapses occurring in larger, mushroom shaped spines [2]. Stronger synapses are defined relatively, as synapses that contribute an increased depolarization to a neuronal membrane upon activation, and are thus more likely to generate an action potential in their postsynaptic neuron. Long-lasting enhancement of synaptic strength elicited by specific patterns of high frequency (Tetanic) synaptic stimulation, resulting in addition of new glutamate receptors and enlargement of the synapse, is referred to as long term potentiation (LTP) [3, 4]. Long-lasting suppression of synaptic strength elicited by specific patterns of long term, low frequency synaptic stimulation, resulting in removal of glutamate receptors

and shrinkage of the synapse, is referred to as long term depression (LTD) [5-8]. Although LTP and LTD have opposing effects on the strength of the synapse, they are both controlled by the precise pattern of Ca^{2+} influx through N-methyl-D-aspartate-type glutamate receptors (NMDAR) [3, 9]. High frequency, larger amplitude Ca^{2+} influx through the NMDAR into the postsynaptic neuron promotes the formation of LTP, whereas lower frequency, lower amplitude, and prolonged Ca^{2+} influx through the NMDAR promotes LTD [10, 11].

The Postsynaptic Density

Function of the Postsynaptic Density

The postsynaptic density (PSD) is an electron-dense region in the spines of excitatory postsynaptic neurons that contains a high concentration of glutamate receptors, cytoskeletal proteins, and associated signaling enzymes [9, 12]. The PSD was initially discovered by electron microscopists as a densely staining slab of proteinaceous material attached to the face of excitatory postsynaptic neurons adjoining presynaptic vesicle release sites [13-16]. Investigation of its biochemical composition was facilitated by the isolation of a specific subcellular fraction from synaptosomes through the use of differential detergent extraction [17, 18]. The PSD forms a dense and highly dynamic scaffold that underlies the spatial organization of signaling events that serve to regulate the strength of the synapse [9, 19].

Glutamate Receptors in the Postsynaptic Density

The postsynaptic membrane in spine synapses contains two distinct types of ligand-gated ion channels that are receptors for the neurotransmitter glutamate: 2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid receptors (AMPAR) and N-methyl-D-aspartate receptors (NMDAR). When bound to glutamate, AMPARs produce a predictable rapid, transient depolarization of the spine membrane, resulting from flux of K^+ and Na^+ ions [20]. Unlike AMPARs, binding of glutamate is not sufficient to cause opening of NMDARs due to a bound Mg^{2+} ion that blocks ion flow through the channel pore. Blockage of the NMDAR can only be relieved when glutamate release occurs concomitantly with a strong depolarization of the postsynaptic membrane (caused, for example, by a back propagating action potential) [21-24]. NMDARs are permeable to Ca^{2+} in addition to Na^+ and K^+; thus, opening of an NMDAR channel causes an influx of Ca^{2+} into the cytosolic region of the spine [25]. As stated above, the precise pattern of Ca^{2+} influx through NMDARs promotes enhancement or suppression of synaptic strength, primarily through the regulation of spine shape, modulation of channel conductance, or insertion/deletion of AMPARs from the synapse. NMDARs also contain long carboxyl terminal tail regions in excess of 600 amino acids in length that extend into the cytosol of the spine and form the basis of the scaffolding structure that organizes cell signaling machinery in the PSD [26].

Scaffolding Molecules in the Postsynaptic Density

Three major classes of scaffolding molecules function to organize signaling enzymes and glutamate receptors in the PSD: PSD-95, or membrane-associated guanylate

kinases (MAGUK); SH3 domain and ankyrin repeat domain proteins (SHANK), or proline-rich-synapse-associated proteins (ProSAP); and Homers.

The four members of the MAGUK class of scaffolds (PSD-95/SAP-90, PSD-93/chapsyn-110, SAP-97, and SAP-102) are located in close proximity to the postsynaptic membrane (12-20 nm) and serve to physically link AMPARs and NMDARs to signaling molecules in the PSD [27, 28]. Each MAGUK contains multiple protein docking sites arranged in a conserved domain architecture consisting of three PDZ domains, an SH3 domain, and a catalytically inactive guanylate kinase (GuK) domain [29]. Because of PSD-95s critical role in receptor clustering, abundance, and protein localization at the synapse, it is the most well characterized PDZ domain-containing protein [9, 19, 30, 31]. It has been shown to form multimers, mediated by head to head interactions via its amino termini [32, 33]. The first two PDZ domains of PSD-95 have been shown to interact with several synaptic membrane proteins, including the tails of the GluN1 and GluN2 subunits of the NMDAR, transmembrane AMPAR regulatory proteins (TARPs) (also known as Stargazins), and Neuroligin [34-37]. The PDZ domains in PSD-95 have also been shown to associate with the cytosolic signaling enzymes neuronal nitric oxide synthase (nNOS) and synaptic GTPase activating protein (synGAP) [38, 39]. The SH3 domain of PSD-95 can bind directly to A-kinase anchoring protein (AKAP), a scaffold protein that serves to position both protein kinase C (PKC), and the protein phosphatase Calcineurin (PP2B) in the PSD [6, 40, 41]. The GuK domain of PSD-95 can be linked to the SHANK family of PSD scaffolds by its interaction with guanylate kinase-associating proteins (GKAPs) [42-44].

The SHANK class of scaffolds (SHANK1-3) is located slightly deeper into the PSD than MAGUKs (~25 nm from the postsynaptic membrane) and serves as a "Scaffold of scaffolds" [45]. The core domain structure of full length members of the SHANK family consists of ankyrin repeats, followed by an SH3 domain, a PDZ domain, a proline rich region, and a sterile alpha motif (SAM) [45]. Genes coding for SHANKs undergo alternative splicing near domain junctions, producing considerable variation in the functionality of each protein [44, 46, 47]. As mentioned above, SHANKs are linked to MAGUK scaffolds through interactions with the GKAP family of proteins. This association is mediated via binding of the C-terminal QTRL sequences of GKAPs to the PDZ domain in SHANKs and of the GuK domain of PSD-95 to a region of GKAPs [42-44]. The proline rich domain in SHANK associates with the SH3 domain of Cortactin, an F-actin binding protein, and serves to link SHANK to the actin cytoskeleton [48].

The proline rich region of SHANK scaffolds also serves to link SHANK with the Homer family of PSD scaffolds [49]. Members of the Homer family consist of an Ena-Vasp-Homology (EVH1) domain and a coiled-coil domain that contains several leucine-zipper motifs. The leucine-zipper motifs in the coiled-coil domains of Homer act as an association domain to force its tetramerization [50], a process that exposes its four EVH1 domains for ligand binding. The exposed EVH1 domains bind to proline rich sites in SHANKs [49] and link Homer to the tails of metabotropic glutamate receptors (mGluRs) and inositol tri-phosphate (IP3) receptors [51, 52].

Structure and Function of PDZ Domains

PDZ domains are ubiquitous, small (~90 residue), compact, modular protein-binding domains that hold together and organize membrane-associated signal transduction complexes [29, 34, 53, 54]. As mentioned above, these domains are often found in multi-domain scaffolding proteins (e.g. PSD-95, SHANK) that link together large molecular complexes at specific locations within the PSD. PDZ domains contain a ligand binding pocket that binds short peptide motifs at the C-termini of partner proteins [55, 56]. Some PDZ domains have also been shown to bind phospholipids and other PDZ domains [38, 57, 58]. Based on their specificity for C-terminal peptide sequences, PDZ domains are divided into three distinct classes [53, 56]. Type I PDZ domains bind C-terminal X-S/T-X-V/I/ L peptide motifs, where X denotes any amino acid. Type II PDZ domains bind C-terminal X-Φ-X-Φ peptide motifs, where Φ denotes the bulky, hydrophobic residues V/Y/F/L/I. Type III PDZ domains bind C-terminal X-D/E-X-V/ L peptide motifs. The affinities of PDZ domains for their ligands range from 0.1 to 10 μM [56, 59-62].

PDZ domains consist of a five-stranded antiparallel β-barrel flanked by three α-helices (Fig. 1.1A), with a ~700 $Å^3$ hydrophobic ligand binding pocket formed by the β2 strand, β1- β2 loop and side chains from the β2 strand and α2 helix [61]. Recognition of the C-terminal carboxylate groups of peptide ligands is performed by main chain amides provided by the β1- β2 loop (also known as the carboxylate binding or GLGF loop) and a highly ordered water molecule constrained by a partially buried arginine side chain in the interior of the ligand binding pocket [55]. Selective recognition of peptide ligands is

provided by van der Waals contacts between the hydrophobic pocket and the side chain of the C-terminal ligand residue, and by the formation of hydrogen bonds between the GLGF loop and terminal carboxylate group of the ligand. Stabilization of the extended peptide ligand in the binding pocket is provided by hydrogen bonds between main chain amido nitrogens and carbonyls located in both the β2 strand and peptide ligand. Very little conformational change is associated with peptide binding, other than a slight shift of the GLGF loop towards the carboxylate group of the peptide ligand (Fig. 1.1B). For example, based on a comparison of all α-carbons in the PDZ domain fold, the ligand bound and free forms of the PDZ3 domain from PSD-95 differ only by an RMSD of 0.9 Å (Fig. 1.1B).

In addition to binding C-terminal peptide motifs, certain PDZ domains, such as the PDZ domain of nNOS, can bind homotypically to other PDZ domains via a β-hairpin that immediately follows the core domain. The β-hairpin structure of nNOS binds to the ligand binding pocket of PDZ2 from PSD-95 or to the PDZ domain of syntrophin to form a complex in which the binding pocket of the nNOS PDZ domain (PDZbh) remains unoccupied and capable of binding to a PDZ domain ligand from another protein [63, 64] (Fig. 1.1C and D).

Because of the specificity and high affinity of PDZ domains for their C-terminal ligands and for PDZbh domains, we reasoned that these interacting pairs could provide the core components of an affinity chromatography system, including affinity resins, affinity tags, and elution agents. In Chapter 2 we show that affinity columns containing the PDZ domains of PSD-95 can be used to purify active PDZ domain-binding proteins to very

high purity in a single step. Solid supports derivatized with recombinant PDZ domains from PSD-95 were prepared and used to purify five heterologously expressed neuronal proteins (GluN2B Tail, synGAP, nNOS PDZ, CRIPT, and cypin) that contain endogenous PDZ domain ligands, with synthetic peptides having the sequences of cognate PDZ domain ligands used as elution agents. We also show that conjugation of PDZ domain-related affinity tags to Proteins Of Interest (POIs) that do not contain endogenous PDZ domains or ligands enables purification of the POIs on PDZ domain-related affinity resins. We have verified that the affinity tags do not alter protein activity by fusing PDZ domain-derived affinity tags to DasherGFP, LacZ, and chloramphenicol acetyltransferase (CAT), purifying the tagged proteins, and assaying their activity using standardized spectrophotometric and fluorescence assays.

Ras and Rap Signaling in the NMDA Receptor Associated Signaling Complex

Influx of Ca^{2+} through the NMDAR leads to the activation of the small GTPases Ras and Rap. Differential activation of Ras and Rap acts to modulate synaptic strength by promoting the insertion (exocytosis) or removal (endocytosis), respectively, of AMPARs from the synapse [65] (Fig. 1.2). Ca^{2+} influx through activated NMDARs results in increased levels of cytosolic Ca^{2+} that can bind to several effector molecules within the spine, the most prominent being calmodulin (CaM), a ubiquitous Ca^{2+} binding protein [66-68]. Ca^{2+} bound CaM (Ca^{2+}/CaM) can bind to and activate a multitude of postsynaptic signaling molecules, including Ras-guanine nucleotide-releasing factor 1 (RasGRF1). RasGRF1 catalyzes the activation of Ras by exchanging its bound GDP for GTP [69],

which in turn activates the kinases Raf-1 and B-Raf. Both Raf-1 and B-Raf are capable of phosphorylating and activating the kinase MEK1/2, which in turn activates ERK1/2. Activation of ERK1/2 has been shown to stimulate the insertion of AMPARs at the synapse [65, 70, 71]. Ca^{2+} influx through NMDARs can also induce the synthesis of cAMP by the Ca^{2+}/CaM activated adenylate cyclases AC1 and AC8, resulting in the activation of PKA and subsequently Rap [72-75]. While active Ras and Rap both activate B-Raf, active Rap inhibits the downstream Ras effector, Raf-1. Both Raf-1 and B-Raf activate the ERK pathway, but B-Raf also activates p38 mitogen-activated protein kinase (p38 MAPK) [76]. Therefore, an increase in the proportion of active Rap over Ras can lead to a relative increase in active p38 MAPK [65]. Active p38 MAPK has been shown to activate the small GTPase Rab5, which in turn drives the endocytosis of AMPARs from the synapse [77, 78].

SynGAP (from Synaptic GTPase activating protein) regulates AMPAR levels by catalyzing the inactivation of GTP bound (active) Ras or Rap [39, 79, 80]. synGAP is positioned in close proximity to the cytoplasmic tail regions of the NMDAR through its association with the PDZ domains of PSD-95 [39, 80]. Ca^{2+}/CaM produced by Ca^{2+} influx through activated NMDARs can bind to and activate the prominent postsynaptic protein kinase, Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII), resulting in the rapid phosphorylation of synGAP and alteration of its GAP activity [81]. Activated cyclin-dependent kinase 5 (CDK5), a known binding partner of CaMKII [82], can also modulate synGAPs activity by phosphorylation (L.Washburn, Unpublished Observation). Modulation of synGAPs activity by CaMKII or CDK5 phosphorylation may alter the ratio

of active Ras to Rap in spines, thus pushing the spine towards the insertion (Active Ras) or removal (Active Rap) of AMPARs, subsequently strengthening or weakening the synapse.

Genetic Studies of synGAP Function

Phenotypes of Homo- and Heterozygous synGAP Knockout Mice.

Homozygous deletion of synGAP (synGAP -/-) is lethal in mice [83, 84]; however, the neurons from mutant embryos can be grown in culture and used to study the effects of synGAP deficiency *in vitro* [84]. Cultured synGAP -/- neurons mature significantly faster than wt neurons and ultimately form larger spine heads in mature neurons [84]. Clusters of PSD proteins (PSD-95, NMDAR & AMPAR subunits), normally localized to dendritic shafts in developing neurons, appear in the spine heads of knockout neurons at earlier time points (10 days *in vitro*) and form larger clusters relative to wt neurons. While the majority of wt spines at 10 days *in vitro* still exhibit the appearance of filopodia, a much larger proportion of spines in synGAP -/- neurons have wide heads associated with mature spines and form contacts with presynaptic neurons containing synapsin (a presynaptic marker). These precocious neurons also exhibited an increased frequency and amplitude of miniature excitatory postsynaptic currents (EPSCs), confirming that an increased number of functional synapses were formed relative to wt neurons. Interestingly, the phenotype conferred by the loss of synGAP was not limited to developmental stages, since mature neurons at 21 days *in vitro* possessed larger spine heads and PSD-95 clusters than wt neurons. Effects of the synGAP knockout on spine structure could be partially rescued by transient expression of wild type synGAP at 10

days *in vitro*; however, transfected synGAP isoforms lacking an active RasGAP domain or a C-terminal PDZ domain binding motif failed to rescue the associated phenotype. These results suggest that one regulatory role of synGAP may be to inhibit synapse formation and limit the size of spine heads.

In additional studies of hippocampal neurons cultured from embryos of synGAP -/- mice, mutant neurons exhibited elevated levels of ERK activation and AMPAR insertion at synapses relative to wild type neurons [79, 83, 85, 86]. These data suggest that synGAP regulates AMPAR insertion in spines through its effect on ERK signaling. Regulation of synGAP activity by the NMDAR may be a common pathway for directing changes in synaptic strength by controlling spine morphology and AMPAR insertion/removal in a coordinated fashion.

Mice with heterozygous deletion of synGAP (synGAP -/+) were found to exhibit strongly reduced induction of LTP in the CA1 region of the hippocampus in the absence of any detectable changes in NMDAR mediated currents or basal synaptic transmission [87]. In spite of increased baseline levels of ERK activation in synGAP -/+ neuronal cultures, pharmacological stimulation of the NMDAR still induced a robust activation of ERK. This observation suggested that while synGAP is involved in regulation of the ERK pathway, its role in LTP formation may involve targets downstream of ERK. In keeping with this hypothesis, Komiyama et al. [87] found that the amount of synaptic potentiation induced by LTP stimulation protocols that act independently of ERK, was also reduced in synGAP -/+ neurons. They also observed these same deficiencies in LTP and elevated

levels of active ERK in synGAP -/+ HRas -/- double mutant mice, suggesting that synGAP may regulate additional small GTPases other than Ras. In addition to LTP deficits, synGAP -/+ mice exhibit slightly impaired spatial memory when assessed by the Morris Water Maze Test [87].

Additional work by Carlisle et al. [88] demonstrated that heterozygous deletion of synGAP in mice was sufficient to cause the formation of an excess of mushroom shaped spines in adult brains. Because spine morphology is controlled by remodeling of its underlying actin cytoskeleton, this result indicated that synGAP was involved in the steady-state regulation of actin in mature spines. In addition to an elevation in active Ras levels in synGAP -/+ neurons, the level of active Rac, a known regulator of the actin cytoskeleton, was also elevated. Phosphorylation and subsequent inactivation of the F-actin severing protein Cofilin, a downstream effector of Rac activation, was also upregulated in these mice. Increased inhibition of Cofilin, resulting in a deficiency of actin filament severing activity, could lead to the observed increase in the number of mushroom spines. Treatment of wt neuronal cultures with NMDA resulted in the transient dephosphorylation, and subsequent activation of Cofilin. SynGAP -/+ or -/- neuronal cultures treated with NMDA did not produce this transient dephosphorylation and also showed depressed excitatory postsynaptic potentials (EPSPs) following NMDA application [88]. It is unlikely that the absence of synGAP's RapGAP activity accounted for the increased number of mushroom spines in the mutant mice, because inhibition of the prominent PSD RapGAP SPAR produced the opposite phenotype [89]. The phenotype observed in synGAP mutant spines is more likely due to loss of RasGAP activity, which

can cause an increase in the steady-state levels of active Ras, resulting in the downstream activation of Rac through the RacGEF Tiam1 [90-92].

Mushroom spines exhibit increased stability relative to thin spines [93, 94] and are less responsive to LTP-inducing stimuli [95]. The contrasting properties of mushroom and thin spines have led to the hypothesis that thin spines are "learning" spines and mushroom spines are "memory" spines [2]. The finding that synGAP -/+ mice have a significantly increased number of mushroom spines relative to wt mice suggests that synGAP normally acts to downregulate signaling pathways responsible for the growth of spine heads and may help to maintain spines in a plastic state [88].

Phenotype of synGAP Mutations in Human Subjects

Mutations in the human synGAP-α gene (SYNGAP1) appear to contribute to non-syndromic intellectual disability (NSID) and may be associated with autism spectrum disorders (ASD) [96, 97]. When SYNGAP1 was sequenced in 154 patients with sporadic NSID (no familial history of NSID), five heterozygous *de novo* truncating mutations (K108Vfs25*, K138*, R579*, L813Rfs835*, Q893Rfs22*) were identified in NSID patients without ASD. A heterozygous *de novo* truncating mutation in SYNGAP1 was also identified in a patient suffering from both NSID and ASD (E706Lfs38*). In contrast, no *de novo* truncating mutations in SYNGAP1 were identified in samples from 570 control subjects, 142 ASD subjects, or 143 subjects suffering from schizophrenia [96, 97]. In total, SYNGAP1 appears to be mutated in 4% of sporadic NSID patients. The heterozygous mutations in SYNGAP1 residing on chromosome 6 may be especially

penetrant, since most mutations identified in genes associated with NSID are located on the X-chromosome or are autosomal recessive [98, 99].

The mutations in SYNGAP1 are likely pathogenic because they lack the RasGAP domain and the C-terminal QTRV motif responsible for PDZ domain binding, both of which are essential for synaptic plasticity and spine morphogenesis required for learning and memory [84, 85, 87]. The absence of morphologic, radiologic, and metabolic features associated with syndromic ID are consistent with the fact that synGAP is expressed exclusively in the brain [39, 80]. Mild epilepsy in four NSID patients with synGAP mutations suggests that disruption of synGAP activity may predispose them to seizures by increasing postsynaptic AMPAR recruitment, resulting in increased excitatory synaptic transmission, as described in mice with SYNGAP1 mutations [85, 86]. The fact that epilepsy is well controlled in these patients by topiramate or valproate is consistent with this hypothesis, since both of these drugs inhibit AMPAR activity [100, 101].

Domain Architecture and Structure of synGAP

SynGAP is a dual Ras and Rap GTPase activating protein (GAP) that is highly concentrated in the PSD of excitatory synapses [39, 79, 80]. Its domain structure consists of an N-terminal pleckstrin homology (PH) domain, a C2 domain, a RasGAP domain, an extended disordered region that contains poly-histidine and poly-proline stretches, a coiled-coil motif, and a C-terminal PDZ domain binding ligand (Fig. 1.3). PH domains have been shown to bind phospholipids and proteins, and may function as membrane

recruitment modules [102]. C2 domains have been described as calcium-dependent phospholipid binding modules capable of binding inositol polyphosphates, phosphatidylserine and phosphatidylcholine [103], and have been shown to bind additional ligands ranging from small molecules to proteins [104-106]. In RasGAPs, C2 domains typically occur between an amino terminal PH domain and a C-terminal RasGAP domain [107]; however, little is known about their functional role(s). RasGAP domains are responsible for stimulating the intrinsic GTPase activity of GTP bound Ras, resulting in its inactivation.

In spite of synGAP's homology to the two RasGAPs, p120GAP and neurofibromin, synGAP stimulates the GTPase activity of Rap much more potently than of Ras [79]. A fragment of synGAP containing its C2 and RasGAP domains stimulates intrinsic Rap1 GAP activity by four orders of magnitude, with a k_{cat} of 0.8 s^{-1}, and stimulates the intrinsic HRas GAP activity by two orders of magnitude [108]. The RapGAP activity of synGAP is unique in that it requires the presence of its C2 domain, which makes synGAP the only example of a RapGAP that utilizes a second domain for its catalytic activity, and suggests a new function of C2 domains.

RasGAPs such as p120GAP accelerate hydrolysis of GTP bound to Ras by providing a conserved arginine residue, dubbed the "arginine finger", to stabilize the γ-phosphate leaving group in GTP [109, 110]. Coincident with insertion of the arginine finger, a conserved glutamine in Ras (Q61) coordinates a nucleophilic water molecule required for catalysis. Unlike Ras, Rap contains a threonine residue at position 61 (T61),

which lacks the ability to coordinate the catalytic water molecule. Canonical RapGAPs circumvent the absence of a catalytic glutamine in Rap by providing an asparagine residue dubbed the "asparagine thumb" [111].

A structure of the C2-GAP region of synGAP (Residues 229-725) confirmed the RasGAP fold for synGAPs RasGAP domain and, in conjunction with biochemical assays and homology modeling, suggested a catalytic mechanism similar to canonical RasGAPs but distinct from canonical RapGAPs [108]. A structural alignment of 153 of 207 equivalent Cα positions in synGAPs RasGAP domain with the RasGAP domain in p120GAP (RMSD of 1.2 Å) [112] confirmed the presence of an arginine finger and canonical RasGAP fold in synGAP. While a large portion of the electron density of synGAP's C2 domain was untraceable, the C-terminal and adjacent β-strands were well enough defined that, when coupled with homology modeling, the relative orientation of the C2 domain with the RasGAP domain could be determined.

The results of a structural alignment of the C2-RasGAP domains of synGAP with the RasGAP domain of p120GAP in complex with Ras [109] suggested an approximate model for the Rap-synGAP complex. Rap interacts with the RasGAP domain of synGAP in a manner similar to Ras, which positions synGAP's C2 domain in proximity to the switch II region of Rap that contains T61. This model was supported by the observation that the RasGAP compromising mutations of the catalytic arginine finger to lysine or proline (R470-K/P) [113, 114] reduced synGAP's RapGAP activity by 100-fold (R470K) or rendered it nearly inactive (R470P). Consistent with their model of

catalysis, G12V Rap, which was regulated by canonical RapGAPs containing an asparagine thumb [115], was not sensitive to synGAP. Interestingly, a T61N mutation in Rap did not significantly affect the kinetics of synGAP mediated GAP inactivation of Rap [108]. Based on the aforementioned observations, it appears that the mode of Rap inactivation by synGAP may be distinct from that of other Rap GAPs.

Regulation of synGAP Activity by Phosphorylation

Work by former members of the Kennedy Laboratory, Jeong Oh and Pasquale Manzerra [81], showed that the RasGAP activity of synGAP is regulated by CaMKII. Phosphorylation of synGAP by CaMKII at multiple sites, including S750/S751/S756, S764/S765, S1058, S1093/1095, S1099, and S1123 (Fig. 1.4), increased its RasGAP activity by 70-95%. SynGAP bound to purified PSD fractions and full length recombinant synGAP bound tightly to Hi-5 cell membranes were rapidly phosphorylated with endogenous or exogenous CaMKII, respectively, to a stoichiometry of 4-5 moles of phosphate per mole of synGAP in 2.5 minutes. The rapid phosphorylation of both recombinant, membrane-bound and native PSD-bound synGAP suggested the presence of several high affinity CaMKII sites, and resulted in a 2-fold increase in the RasGAP activity of synGAP. Phosphorylation by CaMKII increased synGAP's activity in a dose-dependent manner, and these effects could be blocked by the addition of inhibiting monoclonal antibodies against CaMKII. Phosphorylation by CaMKII of several synGAP constructs, followed by MALDI-TOF and tandem mass spectrometry, identified multiple synGAP peptides phosphorylated at sites S750/S751/S756, S764/S765, S1093/S1095, S1058, S1099, or S1123. To determine which phosphorylation sites were responsible for

regulating synGAP's activity, a series of mutants containing specific combinations of serine-to-alanine mutations were constructed and expressed in Hi-5 cells. The GAP activity of these membrane bound proteins was measured following phosphorylation by CaMKII. While GAP activity of wild type and S1058A/S1123A synGAP was increased by 70% following phosphorylation by CaMKII, the activity of the S764A/S765A mutant was only increased by 40%. A mutant containing the S1058A/S1062A/S1064A/S1093A/ S1095A/S1123A mutations (an S1099A mutation couldn't be made) was activated by 40% following CaMKII treatment, but activated by only 20% when additional S764A and S765A mutations were incorporated. The remaining 20% activation of synGAP following CaMKII phosphorylation was attributed to phosphorylation at S750/S751/S756 and/or S1099. Interestingly, all CaMKII phosphorylation sites involved in regulation of synGAP's RasGAP activity were located in its disordered region, suggesting a regulatory role for this region.

Phosphosite specific antibodies raised against peptides phosphorylated at sites S765 and S1123 were used to confirm that phosphorylation of S765 and S1123 increased in cortical neurons following a 15 second exposure to the NMDAR agonist, N-methyl-D-aspartate [81]. Proteomic investigations of phosphorylation sites in PSD proteins have also confirmed the presence of phosphorylation of synGAP at S750/S751, S1058, T1062, and S1099 residues [116-118].

Work by a former member of the. Kennedy Laboratory, Lorraine R. Washburn [119], suggested that the GAP activity of synGAP might be differentially regulated by the

protein kinases, CaMKII and CDK5. Phosphorylation of PSD-bound synGAP by endogenous CaMKII increased both the Ras and RapGAP activities of synGAP approximately two-fold, thus accelerating the rates of inactivation of both Ras and Rap proportionately. Phosphorylation by endogenous kinases of PSD-bound synGAP in the absence of Ca^{2+} or CaM (CaMKII independent sites), followed by tandem mass spectrometry, identified a phosphorylation site at S773 of synGAP (Fig. 1.4), located in a CDK5 consensus sequence (S-P-X-R/K). Phosphorylation of PSD bound synGAP with exogenous CDK5 increased the RasGAP activity of synGAP in a dose dependent manner, but left its RapGAP activity unchanged. This result suggested that phosphorylation of synGAP by CDK5 might lead to an alteration in the ratio of active Ras and Rap in spines, thus pushing the spine towards the removal of AMPA receptors (Active Rap), and subsequent weakening of the synapse. CDK5 has previously been implicated in the regulation of synaptic plasticity [120, 121], and has been localized to synapses and shown to interact with or regulate many synaptic proteins, including NR2A [122], PSD-95 [123], and CaMKII [82].

Moving Forward: Studying the Regulation of synGAP Activity by Phosphorylation Using Purified Proteins

To date, all biochemical studies of the regulation of synGAP activity by protein kinases have utilized impure preparations of synGAP embedded in the PSD or bound to the membranes of insect cells [79, 81, 119]. While these studies have provided insight into the regulation of synGAP's activity, one cannot ascertain the exact contributions from additional GTPases, GAPs, GEFS, kinases, phosphatases, and downstream effectors

present in impure synGAP preparations. As described above, the differential activation of Ras and Rap signaling cascades is believed to be partially responsible for determining whether strengthening or weakening of the synapse occurs through exoctytosis or endocytosis of AMPARs [65]. Due to the extensive amount of crosstalk between the Ras and Rap pathways [31] (Fig. 1.2), and the long time scales required for detecting changes in synaptic strength in *in vivo* experiments, the precise roles of the two pathways have been difficult to distinguish with precision [84, 85, 87]. I sought to clarify the effects of phosphorylation of synGAP on its Ras and Rap GAP activities by preparing purified, soluble recombinant synGAP, as well as recombinant isoforms of Ras and Rap (HRas, Rap1, Rap2). I used pure CaMKII, CDK5, PLK2, and CaM for all studies of phosphorylation. I have circumvented the aforementioned issues and have quantitatively assessed the specific effects of phosphorylation by CaMKII, CDK5, and PLK2, in the presence and absence of Ca^{2+}/CaM, on synGAP activity. Using mass spectrometry, we have confirmed the presence of previously identified CaMKII and CDK5 sites in synGAP [81, 119], and have identified novel sites of phosphorylation by CaMKII, CDK5, and PLK2. We have shown that the net effect of phosphorylation of synGAP by CaMKII, CDK5, and PLK2 is an increase in its GAP activity toward HRas and Rap1. In contrast, there is no effect on its GAP activity toward Rap2. Additionally, we have confirmed the effects of specific CDK5 phosphorylation sites on synGAP activity. In the course of this work, we also found, unexpectedly, that synGAP is itself a Ca^{2+}/CaM binding protein. While Ca^{2+}/CaM binding does not directly affect synGAP activity, it causes a conformational change in synGAP that increases the rate of its phosphorylation and

exposes additional phosphorylation sites that are inaccessible in the absence of Ca^{2+}/CaM.

REFERENCES

[1] T.V. Bliss, T. Lomo, Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. The Journal of physiology 232 (1973) 331-356.

[2] J.N. Bourne, K.M. Harris, Balancing structure and function at hippocampal dendritic spines. Annual review of neuroscience 31 (2008) 47-67.

[3] T.V. Bliss, G.L. Collingridge, A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361 (1993) 31-39.

[4] R.C. Malenka, R.A. Nicoll, Long-term potentiation--a decade of progress? Science (New York, NY) 285 (1999) 1870-1874.

[5] S.M. Dudek, M.F. Bear, Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. Proceedings of the National Academy of Sciences of the United States of America 89 (1992) 4363-4367.

[6] R.M. Mulkey, S. Endo, S. Shenolikar, R.C. Malenka, Involvement of a calcineurin/inhibitor-1 phosphatase cascade in hippocampal long-term depression. Nature 369 (1994) 486-488.

[7] R.M. Mulkey, C.E. Herron, R.C. Malenka, An essential role for protein phosphatases in hippocampal long-term depression. Science (New York, NY) 261 (1993) 1051-1055.

[8] R.M. Mulkey, R.C. Malenka, Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus. Neuron 9 (1992) 967-975.

[9] M.B. Kennedy, Signal-processing machines at the postsynaptic density. Science (New York, NY) 290 (2000) 750-754.

[10] K.M. Franks, T.J. Sejnowski, Complexity of calcium signaling in synaptic spines. BioEssays : news and reviews in molecular, cellular and developmental biology 24 (2002) 1130-1144.

[11] P.J. Sjostrom, S.B. Nelson, Spike timing, calcium signals and synaptic plasticity. Current opinion in neurobiology 12 (2002) 305-314.

[12] M.B. Kennedy, The postsynaptic density at glutamatergic synapses. Trends in neurosciences 20 (1997) 264-268.

[13] E.G. Gray, Axo-somatic and axo-dendritic synapses of the cerebral cortex: an electron microscope study. Journal of anatomy 93 (1959) 420-433.

[14] E.G. Gray, Synaptic fine structure and nuclear, cytoplasmic and extracellular networks: The stereoframework concept. Journal of neurocytology 4 (1975) 315-339.

[15] S.L. Palay, Synapses in the central nervous system. The Journal of biophysical and biochemical cytology 2 (1956) 193-202.

[16] K. Akert, H. Moor, K. Pfenninger, C. Sandri, Contributions of new impregnation methods and freeze etching to the problems of synaptic fine structure. Progress in brain research 31 (1969) 223-240.

[17] F. Blomberg, R.S. Cohen, P. Siekevitz, The structure of postsynaptic densities isolated from dog cerebral cortex. II. Characterization and arrangement of some of the major proteins within the structure. The Journal of cell biology 74 (1977) 204-225.

[18] R.S. Cohen, F. Blomberg, K. Berzins, P. Siekevitz, The structure of postsynaptic densities isolated from dog cerebral cortex. I. Overall morphology and protein composition. The Journal of cell biology 74 (1977) 181-203.

[19] M. Sheng, E. Kim, The postsynaptic organization of synapses. Cold Spring Harbor perspectives in biology 3 (2011).

[20] P.H. Seeburg, The TINS/TiPS Lecture. The molecular biology of mammalian glutamate receptor channels. Trends in neurosciences 16 (1993) 359-365.

[21] N. Spruston, Y. Schiller, G. Stuart, B. Sakmann, Activity-dependent action potential invasion and calcium influx into hippocampal CA1 dendrites. Science (New York, NY) 268 (1995) 297-300.

[22] G. Stuart, N. Spruston, B. Sakmann, M. Hausser, Action potential initiation and backpropagation in neurons of the mammalian CNS. Trends in neurosciences 20 (1997) 125-131.

[23] J.C. Magee, D. Johnston, A synaptically controlled, associative signal for Hebbian plasticity in hippocampal neurons. Science (New York, NY) 275 (1997) 209-213.

[24] J. Magee, D. Hoffman, C. Colbert, D. Johnston, Electrical and calcium signaling in dendrites of hippocampal pyramidal neurons. Annual review of physiology 60 (1998) 327-346.

[25] A.B. MacDermott, M.L. Mayer, G.L. Westbrook, S.J. Smith, J.L. Barker, NMDA-receptor activation increases cytoplasmic calcium concentration in cultured spinal cord neurones. Nature 321 (1986) 519-522.

[26] G. Kohr, V. Jensen, H.J. Koester, A.L. Mihaljevic, J.K. Utvik, A. Kvello, O.P. Ottersen, P.H. Seeburg, R. Sprengel, O. Hvalby, Intracellular domains of NMDA receptor subtypes are determinants for long-term potentiation induction. The Journal of neuroscience : the official journal of the Society for Neuroscience 23 (2003) 10791-10799.

[27] J.D. Petersen, X. Chen, L. Vinade, A. Dosemeci, J.E. Lisman, T.S. Reese, Distribution of postsynaptic density (PSD)-95 and Ca2+/calmodulin-dependent protein kinase II at the PSD. The Journal of neuroscience : the official journal of the Society for Neuroscience 23 (2003) 11270-11278.

[28] J.G. Valtschanoff, R.J. Weinberg, Laminar organization of the NMDA receptor complex within the postsynaptic density. The Journal of neuroscience : the official journal of the Society for Neuroscience 21 (2001) 1211-1217.

[29] H.C. Kornau, P.H. Seeburg, M.B. Kennedy, Interaction of ion channels and receptors with PDZ domain proteins. Current opinion in neurobiology 7 (1997) 368-373.

[30] C.A. Hunt, L.J. Schenker, M.B. Kennedy, PSD-95 is associated with the postsynaptic density and not with the presynaptic membrane at forebrain

synapses. The Journal of neuroscience : the official journal of the Society for Neuroscience 16 (1996) 1380-1388.

[31] M.B. Kennedy, H.C. Beale, H.J. Carlisle, L.R. Washburn, Integration of biochemical signalling in spines. Nature reviews Neuroscience 6 (2005) 423-434.

[32] K.S. Christopherson, N.T. Sweeney, S.E. Craven, R. Kang, D. El-Husseini Ael, D.S. Bredt, Lipid- and protein-mediated multimerization of PSD-95: implications for receptor clustering and assembly of synaptic protein networks. Journal of cell science 116 (2003) 3213-3219.

[33] Y.P. Hsueh, M. Sheng, Requirement of N-terminal cysteines of PSD-95 for PSD-95 multimerization and ternary complex formation, but not for binding to potassium channel Kv1.4. The Journal of biological chemistry 274 (1999) 532-536.

[34] H.C. Kornau, L.T. Schenker, M.B. Kennedy, P.H. Seeburg, Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. Science (New York, NY) 269 (1995) 1737-1740.

[35] M. Niethammer, E. Kim, M. Sheng, Interaction between the C terminus of NMDA receptor subunits and multiple members of the PSD-95 family of membrane-associated guanylate kinases. The Journal of neuroscience : the official journal of the Society for Neuroscience 16 (1996) 2157-2163.

[36] J. Choi, J. Ko, E. Park, J.R. Lee, J. Yoon, S. Lim, E. Kim, Phosphorylation of stargazin by protein kinase A regulates its interaction with PSD-95. The Journal of biological chemistry 277 (2002) 12359-12363.

[37] M. Irie, Y. Hata, M. Takeuchi, K. Ichtchenko, A. Toyoda, K. Hirao, Y. Takai, T.W. Rosahl, T.C. Sudhof, Binding of neuroligins to PSD-95. Science (New York, NY) 277 (1997) 1511-1515.

[38] J.E. Brenman, D.S. Chao, S.H. Gee, A.W. McGee, S.E. Craven, D.R. Santillano, Z. Wu, F. Huang, H. Xia, M.F. Peters, S.C. Froehner, D.S. Bredt, Interaction of nitric oxide synthase with the postsynaptic density protein PSD-95 and alpha1-syntrophin mediated by PDZ domains. Cell 84 (1996) 757-767.

[39] H.J. Chen, M. Rojas-Soto, A. Oguni, M.B. Kennedy, A synaptic Ras-GTPase activating protein (p135 SynGAP) inhibited by CaM kinase II. Neuron 20 (1998) 895-904.

[40] Y. Lin, T. Jover-Mengual, J. Wong, M.V. Bennett, R.S. Zukin, PSD-95 and PKC converge in regulating NMDA receptor trafficking and gating. Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 19902-19907.

[41] K. Kameyama, H.K. Lee, M.F. Bear, R.L. Huganir, Involvement of a postsynaptic protein kinase A substrate in the expression of homosynaptic long-term depression. Neuron 21 (1998) 1163-1175.

[42] T.M. Boeckers, C. Winter, K.H. Smalla, M.R. Kreutz, J. Bockmann, C. Seidenbecher, C.C. Garner, E.D. Gundelfinger, Proline-rich synapse-associated proteins ProSAP1 and ProSAP2 interact with synaptic proteins of the SAPAP/GKAP family. Biochemical and biophysical research communications 264 (1999) 247-252.

[43] S. Naisbitt, E. Kim, R.J. Weinberg, A. Rao, F.C. Yang, A.M. Craig, M. Sheng, Characterization of guanylate kinase-associated protein, a postsynaptic density protein at excitatory synapses that interacts directly with postsynaptic density-95/synapse-associated protein 90. The Journal of neuroscience : the official journal of the Society for Neuroscience 17 (1997) 5687-5696.

[44] I. Yao, Y. Hata, K. Hirao, M. Deguchi, N. Ide, M. Takeuchi, Y. Takai, Synamon, a novel neuronal protein interacting with synapse-associated protein 90/postsynaptic density-95-associated protein. The Journal of biological chemistry 274 (1999) 27463-27466.

[45] M. Sheng, E. Kim, The Shank family of scaffold proteins. Journal of cell science 113 (Pt 11) (2000) 1851-1856.

[46] S. Lim, S. Naisbitt, J. Yoon, J.I. Hwang, P.G. Suh, M. Sheng, E. Kim, Characterization of the Shank family of synaptic proteins. Multiple genes,

alternative splicing, and differential expression in brain and development. The Journal of biological chemistry 274 (1999) 29510-29518.

[47] H. Zitzer, H.H. Honck, D. Bachner, D. Richter, H.J. Kreienkamp, Somatostatin receptor interacting protein defines a novel family of multidomain proteins present in human and rodent brain. The Journal of biological chemistry 274 (1999) 32997-33001.

[48] H. Hering, M. Sheng, Activity-dependent redistribution and essential role of cortactin in dendritic spine morphogenesis. The Journal of neuroscience : the official journal of the Society for Neuroscience 23 (2003) 11759-11769.

[49] J.C. Tu, B. Xiao, S. Naisbitt, J.P. Yuan, R.S. Petralia, P. Brakeman, A. Doan, V.K. Aakalu, A.A. Lanahan, M. Sheng, P.F. Worley, Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of postsynaptic density proteins. Neuron 23 (1999) 583-592.

[50] M.K. Hayashi, H.M. Ames, Y. Hayashi, Tetrameric hub structure of postsynaptic scaffolding protein homer. The Journal of neuroscience : the official journal of the Society for Neuroscience 26 (2006) 8492-8501.

[51] A. Kato, F. Ozawa, Y. Saitoh, K. Hirai, K. Inokuchi, vesl, a gene encoding VASP/Ena family related protein, is upregulated during seizure, long-term potentiation and synaptogenesis. FEBS letters 412 (1997) 183-189.

[52] J.C. Tu, B. Xiao, J.P. Yuan, A.A. Lanahan, K. Leoffert, M. Li, D.J. Linden, P.F. Worley, Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. Neuron 21 (1998) 717-726.

[53] A.Y. Hung, M. Sheng, PDZ domains: structural modules for protein complex assembly. The Journal of biological chemistry 277 (2002) 5699-5702.

[54] K.O. Cho, C.A. Hunt, M.B. Kennedy, The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. Neuron 9 (1992) 929-942.

[55] D.A. Doyle, A. Lee, J. Lewis, E. Kim, M. Sheng, R. MacKinnon, Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 85 (1996) 1067-1076.

[56] Z. Songyang, A.S. Fanning, C. Fu, J. Xu, S.M. Marfatia, A.H. Chishti, A. Crompton, A.C. Chan, J.M. Anderson, L.C. Cantley, Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science (New York, NY) 275 (1997) 73-77.

[57] Y. Ivarsson, A.M. Wawrzyniak, G. Wuytens, M. Kosloff, E. Vermeiren, M. Raport, P. Zimmermann, Cooperative phosphoinositide and peptide binding by PSD-95/discs large/ZO-1 (PDZ) domain of polychaetoid, Drosophila zonulin. The Journal of biological chemistry 286 (2011) 44669-44678.

[58] J.E. Brenman, K.S. Christopherson, S.E. Craven, A.W. McGee, D.S. Bredt, Cloning and characterization of postsynaptic density 93, a nitric oxide synthase interacting protein. The Journal of neuroscience : the official journal of the Society for Neuroscience 16 (1996) 7407-7415.

[59] M. Niethammer, J.G. Valtschanoff, T.M. Kapoor, D.W. Allison, R.J. Weinberg, A.M. Craig, M. Sheng, CRIPT, a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90. Neuron 20 (1998) 693-707.

[60] S.M. Marfatia, J.H. Morais Cabral, L. Lin, C. Hough, P.J. Bryant, L. Stolz, A.H. Chishti, Modular organization of the PDZ domains in the human discs-large protein suggests a mechanism for coupling PDZ domain-binding proteins to ATP and the membrane cytoskeleton. The Journal of cell biology 135 (1996) 753-766.

[61] J.H. Morais Cabral, C. Petosa, M.J. Sutcliffe, S. Raza, O. Byron, F. Poy, S.M. Marfatia, A.H. Chishti, R.C. Liddington, Crystal structure of a PDZ domain. Nature 382 (1996) 649-652.

[62] B.Z. Harris, B.J. Hillier, W.A. Lim, Energetic determinants of internal motif recognition by PDZ domains. Biochemistry 40 (2001) 5921-5930.

[63] B.J. Hillier, K.S. Christopherson, K.E. Prehoda, D.S. Bredt, W.A. Lim, Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. Science (New York, NY) 284 (1999) 812-815.

[64] H. Tochio, Q. Zhang, P. Mandal, M. Li, M. Zhang, Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide. Nature structural biology 6 (1999) 417-421.

[65] J.J. Zhu, Y. Qin, M. Zhao, L. Van Aelst, R. Malinow, Ras and Rap control AMPA receptor trafficking during synaptic plasticity. Cell 110 (2002) 443-455.

[66] D. Chin, A.R. Means, Calmodulin: a prototypical calcium sensor. Trends in cell biology 10 (2000) 322-328.

[67] S. Linse, A. Helmersson, S. Forsen, Calcium binding to calmodulin and its globular domains. The Journal of biological chemistry 266 (1991) 8050-8054.

[68] C.B. Klee, Interaction of CaM with calcium and target proteins., Elsevier, Amsterdam, 1988.

[69] C.L. Farnsworth, N.W. Freshney, L.B. Rosen, A. Ghosh, M.E. Greenberg, L.A. Feig, Calcium activation of Ras mediated by neuronal exchange factor Ras-GRF. Nature 376 (1995) 524-527.

[70] G. Krapivinsky, L. Krapivinsky, Y. Manasian, A. Ivanov, R. Tyzio, C. Pellegrino, Y. Ben-Ari, D.E. Clapham, I. Medina, The NMDA receptor is coupled to the ERK pathway by a direct interaction between NR2B and RasGRF1. Neuron 40 (2003) 775-784.

[71] X. Tian, T. Gotoh, K. Tsuji, E.H. Lo, S. Huang, L.A. Feig, Developmentally regulated role for Ras-GRFs in coupling NMDA glutamate receptors to Ras, Erk and CREB. The EMBO journal 23 (2004) 1567-1575.

[72] G.D. Ferguson, D.R. Storm, Why calcium-stimulated adenylyl cyclases? Physiology 19 (2004) 271-276.

[73] D.M. Chetkovich, J.D. Sweatt, nMDA receptor activation increases cyclic AMP in area CA1 of the hippocampus via calcium/calmodulin stimulation of adenylyl cyclase. Journal of neurochemistry 61 (1993) 1933-1942.

[74] A. Morozov, I.A. Muzzio, R. Bourtchouladze, N. Van-Strien, K. Lapidus, D. Yin, D.G. Winder, J.P. Adams, J.D. Sweatt, E.R. Kandel, Rap1 couples cAMP signaling to a distinct pool of p42/44MAPK regulating excitability, synaptic plasticity, learning, and memory. Neuron 39 (2003) 309-325.

[75] S.S. Grewal, A.M. Horgan, R.D. York, G.S. Withers, G.A. Banker, P.J. Stork, Neuronal calcium activates a Rap1 and B-Raf signaling pathway via the cyclic adenosine monophosphate-dependent protein kinase. The Journal of biological chemistry 275 (2000) 3722-3728.

[76] P.J. Stork, Does Rap1 deserve a bad Rap? Trends in biochemical sciences 28 (2003) 267-275.

[77] T.C. Brown, I.C. Tran, D.S. Backos, J.A. Esteban, NMDA receptor-dependent activation of the small GTPase Rab5 drives the removal of synaptic AMPA receptors during hippocampal LTD. Neuron 45 (2005) 81-94.

[78] C.C. Huang, J.L. You, M.Y. Wu, K.S. Hsu, Rap1-induced p38 mitogen-activated protein kinase activation facilitates AMPA receptor trafficking via the GDI.Rab5 complex. Potential role in (S)-3,5-dihydroxyphenylglycene-induced long term depression. The Journal of biological chemistry 279 (2004) 12286-12292.

[79] G. Krapivinsky, I. Medina, L. Krapivinsky, S. Gapon, D.E. Clapham, SynGAP-MUPP1-CaMKII synaptic complexes regulate p38 MAP kinase activity and NMDA receptor-dependent synaptic AMPA receptor potentiation. Neuron 43 (2004) 563-574.

[80] J.H. Kim, D. Liao, L.F. Lau, R.L. Huganir, SynGAP: a synaptic RasGAP that associates with the PSD-95/SAP90 protein family. Neuron 20 (1998) 683-691.

[81] J.S. Oh, P. Manzerra, M.B. Kennedy, Regulation of the neuron-specific Ras GTPase-activating protein, synGAP, by Ca2+/calmodulin-dependent protein kinase II. The Journal of biological chemistry 279 (2004) 17980-17988.

[82] R. Dhavan, P.L. Greer, M.A. Morabito, L.R. Orlando, L.H. Tsai, The cyclin-dependent kinase 5 activators p35 and p39 interact with the alpha-subunit of Ca2+/calmodulin-dependent protein kinase II and alpha-actinin-1 in a calcium-

dependent manner. The Journal of neuroscience : the official journal of the Society for Neuroscience 22 (2002) 7879-7891.

[83] J.H. Kim, H.K. Lee, K. Takamiya, R.L. Huganir, The role of synaptic GTPase-activating protein in neuronal development and synaptic plasticity. The Journal of neuroscience : the official journal of the Society for Neuroscience 23 (2003) 1119-1124.

[84] L.E. Vazquez, H.J. Chen, I. Sokolova, I. Knuesel, M.B. Kennedy, SynGAP regulates spine formation. The Journal of neuroscience : the official journal of the Society for Neuroscience 24 (2004) 8862-8872.

[85] G. Rumbaugh, J.P. Adams, J.H. Kim, R.L. Huganir, SynGAP regulates synaptic strength and mitogen-activated protein kinases in cultured neurons. Proceedings of the National Academy of Sciences of the United States of America 103 (2006) 4344-4351.

[86] M.J. Kim, A.W. Dunah, Y.T. Wang, M. Sheng, Differential roles of NR2A- and NR2B-containing NMDA receptors in Ras-ERK signaling and AMPA receptor trafficking. Neuron 46 (2005) 745-760.

[87] N.H. Komiyama, A.M. Watabe, H.J. Carlisle, K. Porter, P. Charlesworth, J. Monti, D.J. Strathdee, C.M. O'Carroll, S.J. Martin, R.G. Morris, T.J. O'Dell, S.G. Grant, SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and learning in the complex with postsynaptic density 95 and NMDA receptor. The Journal of neuroscience : the official journal of the Society for Neuroscience 22 (2002) 9721-9732.

[88] H.J. Carlisle, P. Manzerra, E. Marcora, M.B. Kennedy, SynGAP regulates steady-state and activity-dependent phosphorylation of cofilin. The Journal of neuroscience : the official journal of the Society for Neuroscience 28 (2008) 13673-13683.

[89] D.T. Pak, S. Yang, S. Rudolph-Correia, E. Kim, M. Sheng, Regulation of dendritic spine morphology by SPAR, a PSD-95-associated RapGAP. Neuron 31 (2001) 289-303.

[90] K.F. Tolias, J.B. Bikoff, A. Burette, S. Paradis, D. Harrar, S. Tavazoie, R.J. Weinberg, M.E. Greenberg, The Rac1-GEF Tiam1 couples the NMDA receptor to the activity-dependent development of dendritic arbors and spines. Neuron 45 (2005) 525-538.

[91] J.M. Lambert, Q.T. Lambert, G.W. Reuther, A. Malliri, D.P. Siderovski, J. Sondek, J.G. Collard, C.J. Der, Tiam1 mediates Ras activation of Rac by a PI(3)K-independent mechanism. Nature cell biology 4 (2002) 621-625.

[92] I.N. Fleming, I.H. Batty, A.R. Prescott, A. Gray, G.S. Kular, H. Stewart, C.P. Downes, Inositol phospholipids regulate the guanine-nucleotide-exchange factor Tiam1 by facilitating its binding to the plasma membrane and regulating GDP/GTP exchange on Rac1. The Biochemical journal 382 (2004) 857-865.

[93] A.J. Holtmaat, J.T. Trachtenberg, L. Wilbrecht, G.M. Shepherd, X. Zhang, G.W. Knott, K. Svoboda, Transient and persistent dendritic spines in the neocortex in vivo. Neuron 45 (2005) 279-291.

[94] Y. Zuo, A. Lin, P. Chang, W.B. Gan, Development of long-term dendritic spine stability in diverse regions of cerebral cortex. Neuron 46 (2005) 181-189.

[95] M. Matsuzaki, N. Honkura, G.C. Ellis-Davies, H. Kasai, Structural basis of long-term potentiation in single dendritic spines. Nature 429 (2004) 761-766.

[96] F.F. Hamdan, J. Gauthier, D. Spiegelman, A. Noreau, Y. Yang, S. Pellerin, S. Dobrzeniecka, M. Cote, E. Perreau-Linck, L. Carmant, G. D'Anjou, E. Fombonne, A.M. Addington, J.L. Rapoport, L.E. Delisi, M.O. Krebs, F. Mouaffak, R. Joober, L. Mottron, P. Drapeau, C. Marineau, R.G. Lafreniere, J.C. Lacaille, G.A. Rouleau, J.L. Michaud, G. Synapse to Disease, Mutations in SYNGAP1 in autosomal nonsyndromic mental retardation. The New England journal of medicine 360 (2009) 599-605.

[97] F.F. Hamdan, H. Daoud, A. Piton, J. Gauthier, S. Dobrzeniecka, M.O. Krebs, R. Joober, J.C. Lacaille, A. Nadeau, J.M. Milunsky, Z. Wang, L. Carmant, L. Mottron, M.H. Beauchamp, G.A. Rouleau, J.L. Michaud, De novo SYNGAP1

mutations in nonsyndromic intellectual disability and autism. Biological psychiatry 69 (2011) 898-901.

[98] J. Chelly, M. Khelfaoui, F. Francis, B. Cherif, T. Bienvenu, Genetics and pathophysiology of mental retardation. European journal of human genetics : EJHG 14 (2006) 701-713.

[99] J. Gecz, C. Shoubridge, M. Corbett, The genetic landscape of intellectual disability arising from chromosome X. Trends in genetics : TIG 25 (2009) 308-316.

[100] M. Angehagen, L. Ronnback, E. Hansson, E. Ben-Menachem, Topiramate reduces AMPA-induced Ca(2+) transients and inhibits GluR1 subunit phosphorylation in astrocytes from primary cultures. Journal of neurochemistry 94 (2005) 1124-1130.

[101] J. Du, N.A. Gray, C.A. Falke, W. Chen, P. Yuan, S.T. Szabo, H. Einat, H.K. Manji, Modulation of synaptic plasticity by antimanic agents: the role of AMPA glutamate receptor subunit 1 synaptic expression. The Journal of neuroscience : the official journal of the Society for Neuroscience 24 (2004) 6578-6589.

[102] M.A. Lemmon, Pleckstrin homology domains: not just for phosphoinositides. Biochemical Society transactions 32 (2004) 707-711.

[103] E.A. Nalefski, J.J. Falke, The C2 domain calcium-binding motif: structural and functional diversity. Protein science : a publication of the Protein Society 5 (1996) 2375-2390.

[104] J. Rizo, T.C. Sudhof, C2-domains, structure and function of a universal Ca2+-binding domain. The Journal of biological chemistry 273 (1998) 15879-15882.

[105] W. Cho, Membrane targeting by C1 and C2 domains. The Journal of biological chemistry 276 (2001) 32407-32410.

[106] C.H. Benes, N. Wu, A.E. Elia, T. Dharia, L.C. Cantley, S.P. Soltoff, The C2 domain of PKCdelta is a phosphotyrosine binding domain. Cell 121 (2005) 271-280.

[107] A. Bernards, GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and Drosophila. Biochimica et biophysica acta 1603 (2003) 47-82.

[108] V. Pena, M. Hothorn, A. Eberth, N. Kaschau, A. Parret, L. Gremer, F. Bonneau, M.R. Ahmadian, K. Scheffzek, The C2 domain of SynGAP is essential for stimulation of the Rap GTPase reaction. EMBO reports 9 (2008) 350-355.

[109] K. Scheffzek, M.R. Ahmadian, W. Kabsch, L. Wiesmuller, A. Lautwein, F. Schmitz, A. Wittinghofer, The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. Science (New York, NY) 277 (1997) 333-338.

[110] K. Scheffzek, M.R. Ahmadian, L. Wiesmuller, W. Kabsch, P. Stege, F. Schmitz, A. Wittinghofer, Structural analysis of the GAP-related domain from neurofibromin and its implications. The EMBO journal 17 (1998) 4313-4327.

[111] A. Scrima, C. Thomas, D. Deaconescu, A. Wittinghofer, The Rap-RapGAP complex: GTP hydrolysis without catalytic glutamine and arginine residues. The EMBO journal 27 (2008) 1145-1153.

[112] K. Scheffzek, A. Lautwein, W. Kabsch, M.R. Ahmadian, A. Wittinghofer, Crystal structure of the GTPase-activating domain of human p120GAP and implications for the interaction with Ras. Nature 384 (1996) 591-596.

[113] M.R. Ahmadian, P. Stege, K. Scheffzek, A. Wittinghofer, Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. Nature structural biology 4 (1997) 686-689.

[114] A. Klose, M.R. Ahmadian, M. Schuelke, K. Scheffzek, S. Hoffmeyer, A. Gewies, F. Schmitz, D. Kaufmann, H. Peters, A. Wittinghofer, P. Nurnberg, Selective disactivation of neurofibromin GAP activity in neurofibromatosis type 1. Human molecular genetics 7 (1998) 1261-1268.

[115] T. Brinkmann, O. Daumke, U. Herbrand, D. Kuhlmann, P. Stege, M.R. Ahmadian, A. Wittinghofer, Rap-specific GTPase activating protein follows an

alternative mechanism. The Journal of biological chemistry 277 (2002) 12525-12531.

[116] A. Dosemeci, H. Jaffe, Regulation of phosphorylation at the postsynaptic density during different activity states of Ca2+/calmodulin-dependent protein kinase II. Biochemical and biophysical research communications 391 (2010) 78-84.

[117] H. Jaffe, L. Vinade, A. Dosemeci, Identification of novel phosphorylation sites on postsynaptic density proteins. Biochemical and biophysical research communications 321 (2004) 210-218.

[118] J.C. Trinidad, C.G. Specht, A. Thalhammer, R. Schoepfer, A.L. Burlingame, Comprehensive identification of phosphorylation sites in postsynaptic density preparations. Molecular & cellular proteomics : MCP 5 (2006) 914-922.

[119] L.R. Washburn, (Unpublished Observation).

[120] M. Angelo, F. Plattner, K.P. Giese, Cyclin-dependent kinase 5 in synaptic plasticity, learning and memory. Journal of neurochemistry 99 (2006) 353-370.

[121] Z.H. Cheung, A.K. Fu, N.Y. Ip, Synaptic roles of Cdk5: implications in higher cognitive functions and neurodegenerative diseases. Neuron 50 (2006) 13-18.

[122] B.S. Li, M.K. Sun, L. Zhang, S. Takahashi, W. Ma, L. Vinade, A.B. Kulkarni, R.O. Brady, H.C. Pant, Regulation of NMDA receptors by cyclin-dependent kinase-5. Proceedings of the National Academy of Sciences of the United States of America 98 (2001) 12742-12747.

[123] M.A. Morabito, M. Sheng, L.H. Tsai, Cyclin-dependent kinase 5 phosphorylates the N-terminal domain of the postsynaptic density protein PSD-95 in neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 24 (2004) 865-876.

FIGURES

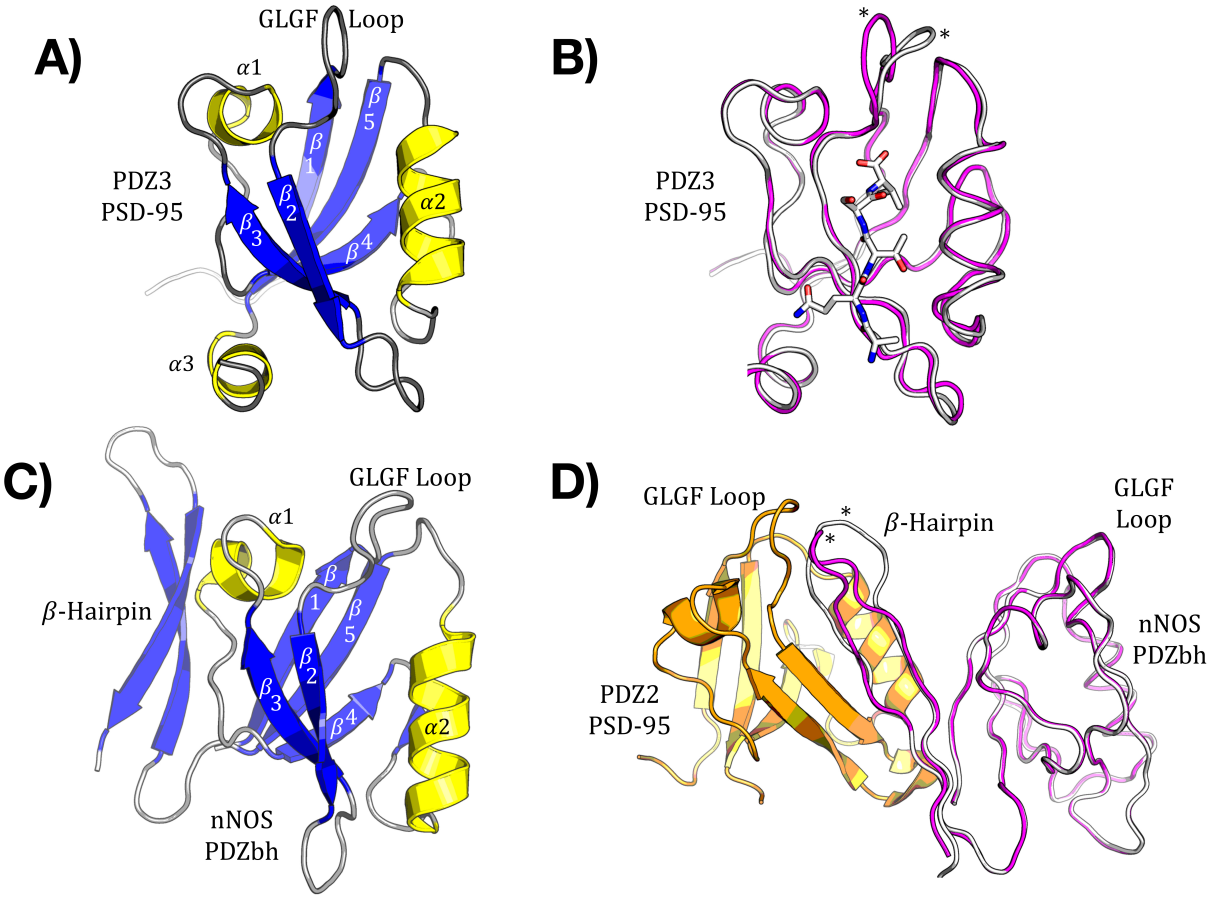


Figure 1.1. Structures of PDZ Domains in Isolation and Complexed With Ligands. **(A)** Ribbon diagram of the three dimensional structure of the PDZ3 domain from PSD-95 (RCSB ID 1BFE). The PDZ domain consists of five β-strands arranged in an antiparallel β-barrel flanked by three α-helices. The PDZ domain ligand inserts in a hydrophobic pocket lined by the β2 strand, β1- β2 loop, and the α2 helix and forms an antiparallel β-sheet with the β2 strand. The β1- β2 loop is required for recognition of the C-terminal carboxylate group of the ligand, and is referred to as the carboxylate-binding loop or GLGF loop. **(B)** Comparison of the ligand-bound (grey) and ligand-free (magenta) forms of the PDZ3 domain from PSD-95. The protein backbones of the ligand-bound (grey, α-carbon traces; RCSB ID 1BE9) and uncomplexed (magenta, α-carbon traces; RCSB ID 1BFE) PDZ3 domains were structurally aligned in UCSF Chimera, producing an RMSD of 0.9 Å. Upon complexation with peptide ligand, S320 in the GLGF loop (marked with an asterisk, *) shifts approximately 4.8 Å. Bound CRIPT peptide (KQTSV) is shown in stick form. **(C)** Ribbon diagram of the three dimensional structure of the PDZ-Beta Hairpin (PDZbh) domain from nNOS (RCSB ID 1QAU). The PDZbh has a polarized

structure with two distinct faces: a ligand binding pocket and β-hairpin face. **(D)** Comparison of the complexed (grey) and free (magenta) forms of the PDZbh domain from nNOS. The protein backbones of uncomplexed PDZbh (RCSB ID 1QAU) and PDZbh bound to PDZ2 of PSD-95 (orange ribbon; RCSB ID 1QAV) were structurally aligned in UCSF Chimera, producing an RMSD of 0.9 Å (0.21 Å RMSD of the bh domain is excluded from the alignment). Upon complexation with PDZ2, D114 in the PDZbh domain (marked with an asterisk, *) shifts approximately 8.5 Å.

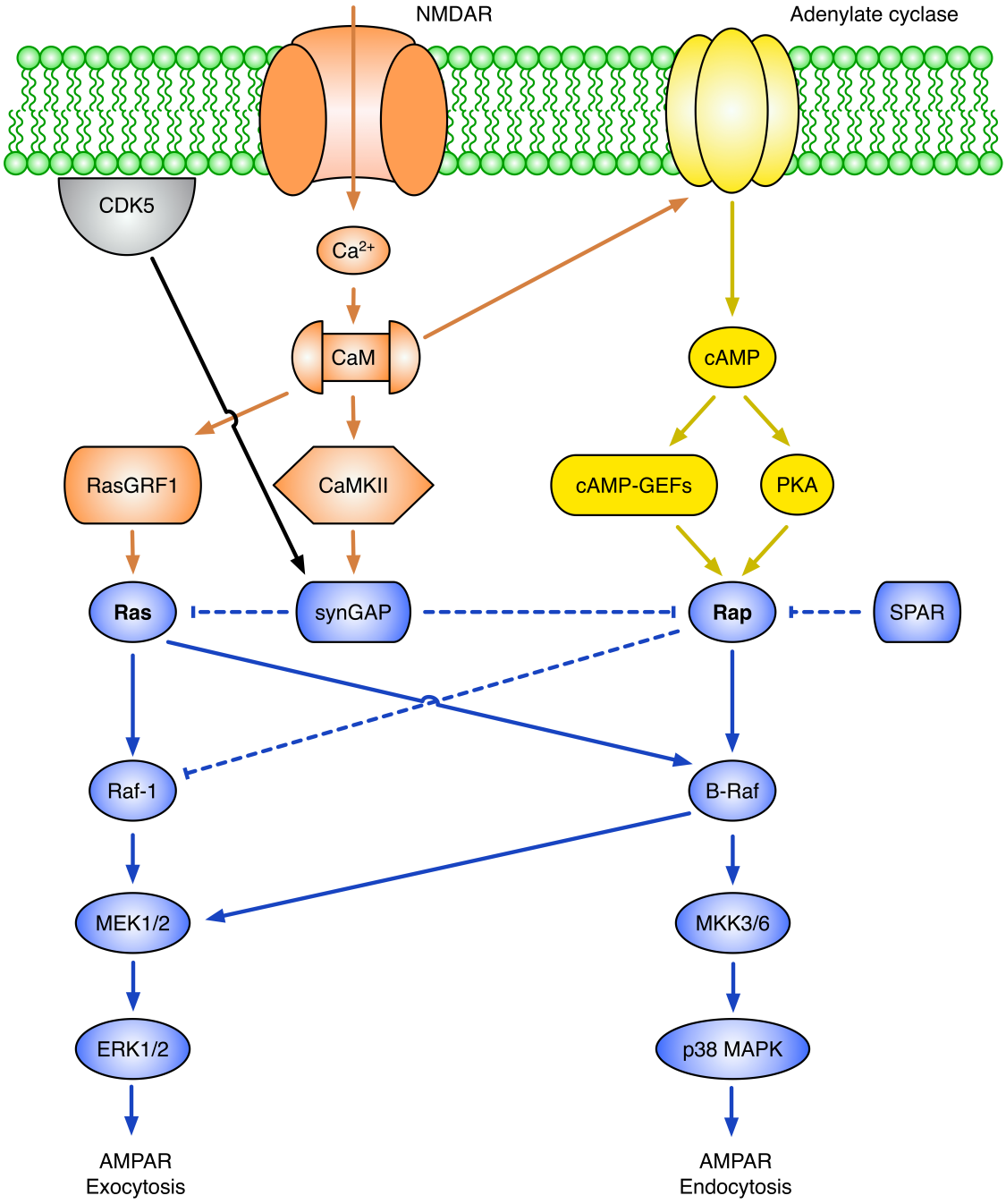


Figure 1.2. Signaling from the NMDAR to Ras and Rap in the Postsynaptic Density. As discussed in the text, Ca^{2+} flowing through the NMDAR regulates signaling pathways that influence the ratio of activated Ras and Rap in the spine. Signaling pathways initiated by the second messengers Ca^{2+} and cAMP are shown in orange and yellow, respectively. Signaling pathways initiated by the small GTPases Ras and Rap are shown in blue.

Signaling pathways with unknown conditions for initiation are shown in black. Small GTPases are highlighted in bold. Solid lines with arrows indicate activation. Dashed lines with a bar indicate inhibition. Quantitative distinctions of pathway activation or inhibition have not been made, since the magnitudes of such influences will vary under different physiological circumstances. Abbreviations: v-raf murine sarcoma viral oncogene homologue B1 (B-Raf); calmodulin (CaM); Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII); cyclic AMP guanine nucleotide exchange factor (cAMP-GEF); cyclin-dependent kinase 5 (CDK5); extracellular signal-regulated kinase 1/2 (ERK1/2); MAPK/ERK Kinase 1/2 (MEK1/2); mitogen-activated protein kinase kinase 3/6 (MKK3/6); p38 mitogen-activated protein kinase (p38 MAPK); N-methyl-D-aspartate Receptor (NMDAR); protein kinase A (PKA); v-raf-1 leukemia viral oncogene 1 (Raf-1); Ras-related protein Rap (Rap); p21Ras (Ras); Ras-Guanine nucleotide-releasing factor 1 (RasGRF1); spine-associated Rap GTPase-activating protein (SPAR); synaptic GTPase activating protein (synGAP). This figure was adapted from Kennedy et al., 2005 [31].

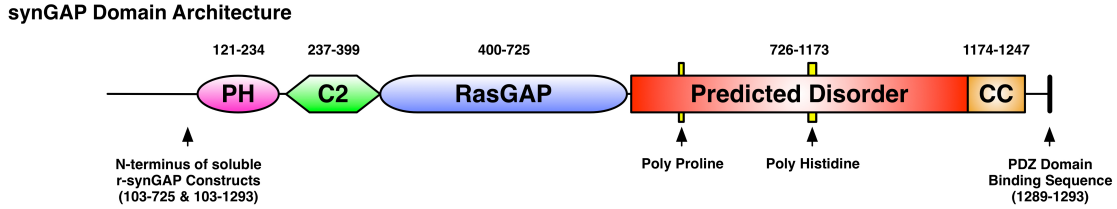


Figure 1.3. Domain Architecture of synGAP-α. Abbreviations: pleckstrin homology domain (PH); C2 domain (C2); Ras GTPase activating protein domain (RasGAP); domain predicted to be disordered by PONDR and DISOPRED algorithms (Predicted Disorder); coiled-coil domain (CC). Also shown: poly proline stretch (P_6); poly histidine stretch (H_{10}); the N-terminus of soluble r-synGAP constructs; the PDZ domain binding sequence.

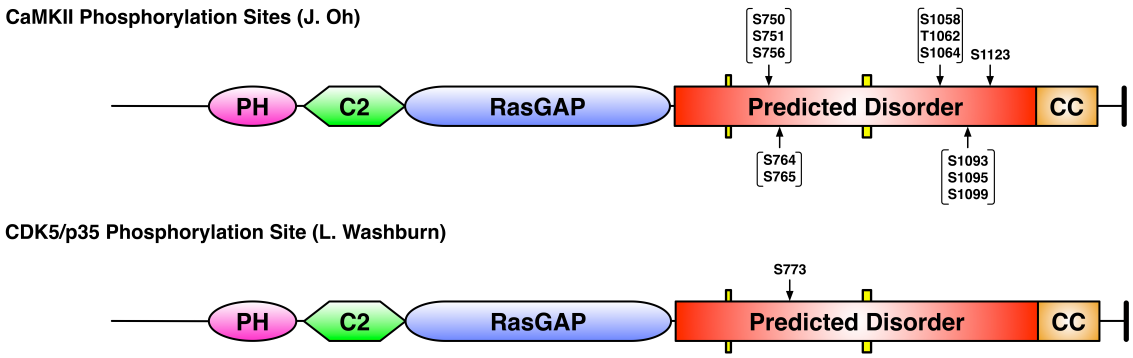


Figure 1.4. Known CaMKII and CDK5 Phosphorylation Sites in synGAP. Top: CaMKII phosphorylation sites in synGAP identified by Jeong Oh et al. [81] that have been shown to increase synGAPs Ras and Rap GAP activity by 2-fold. **Bottom:** CDK5 phosphorylation site in synGAP identified by Lori Washburn (Unpublished Observation) that has been shown to increase synGAPs RasGAP activity by 2-fold. Bracketed residues highlight phosphorylation sites located in single tryptic phosphopeptides of synGAP. The original data was insufficient to assign the precise location of phosphorylation sites.

Chapter II

PDZ AFFINITY CHROMATOGRAPHY: A GENERAL METHOD FOR AFFINITY PURIFICATION OF PROTEINS BASED ON PDZ DOMAINS AND THEIR LIGANDS

AUTHORS

Ward G. Walkup IV* and Mary B. Kennedy

Division of Biology and Biological Engineering, California Institute of Technology, 1200 East California Blvd, Mail Code 216-76, Pasadena CA 91125, USA

*To whom correspondence should be addressed. Tel: +1 6263953924; Fax: +1 6263958474; Email: wwalkup@caltech.edu

Key Words:

HaloTag, PDZ Domain, Affinity Chromatography, NMDA Receptor, synGAP, Neuronal

Author Contributions:

WGWIV conceived, designed, and carried out experiments; analyzed and interpreted data; and wrote the paper. MBK interpreted data and wrote the paper.

ABSTRACT

PDZ (PSD-95, DiscsLarge, ZO1) domains function in nature as protein binding domains within scaffold and membrane-associated proteins. They comprise ~ 90 residues and make specific, high affinity interactions with complementary C-terminal peptide sequences, with other PDZ domains, and with phospholipids. We hypothesized that the specific, strong interactions of PDZ domains with their ligands would make them well suited for use in affinity chromatography. Here we describe a novel affinity chromatography method applicable for the purification of proteins that contain PDZ domain-binding ligands, either naturally or introduced by genetic engineering. We created a series of affinity resins comprised of PDZ domains from the scaffold protein PSD-95, or from neuronal nitric oxide synthase (nNOS), coupled to solid supports. We used them to purify heterologously expressed neuronal proteins or protein domains containing endogenous PDZ domain ligands, eluting the proteins with free PDZ domain peptide ligands. We show that Proteins of Interest (POIs) lacking endogenous PDZ domain ligands can be engineered as fusion products containing C-terminal PDZ domain ligand peptides or internal, N- or C-terminal PDZ domains and then can be purified by the same method. Using this method, we recovered recombinant GFP fused to a PDZ-domain ligand in active form as verified by fluorescence yield. Similarly, chloramphenicol acetyltransferase (CAT) and β-Galactosidase (LacZ) fused to a C-terminal PDZ domain ligand or an N-terminal PDZ domain were purified in active form as assessed by enzymatic assay. In general, PDZ domains and ligands derived from PSD-95 were superior to those from nNOS for this method. PDZ Domain Affinity Chromatography

promises to be a versatile and effective method for purification of a wide variety of natural and recombinant proteins.

INTRODUCTION

PDZ domains are ubiquitous, small (~90 residue), compact, modular protein-binding domains that hold together and organize membrane-associated signal transduction complexes [1-4]. They are often found in multi-domain scaffolding proteins that link together large molecular complexes at specific locations within cells. PDZ domains contain a ligand binding pocket that binds short peptide motifs at the C-termini of partner proteins [5, 6]. Some have also been shown to bind phospholipids and other PDZ domains [7, 8]. They can be divided into three distinct classes based on their specificity for distinct C-terminal peptide sequences [3, 5]. Type I PDZ domains bind C-terminal X-S/T-X-V/I/L peptide motifs, where X denotes any amino acid. Type II PDZ domains bind C-terminal X-Φ-X-Φ peptide motifs, where Φ denotes the bulky, hydrophobic residues V/Y/F/L/I. Type III PDZ domains bind C-terminal X-D/E-X-V/ L peptide motifs. The affinities of PDZ domains for their ligands range from 0.1 to 10 μM [5, 9-12].

In addition to binding C-terminal peptide motifs, certain PDZ domains, such as the PDZ domain of neuronal nitric oxide synthase (nNOS), can bind homotypically to other PDZ domains via a β-Hairpin that immediately follows the core domain. The β-hairpin structure of nNOS binds to the ligand binding pocket of PDZ2 from PSD-95 or to the PDZ domain of syntrophin to form a complex in which the binding pocket of the nNOS PDZ

domain remains unoccupied and capable of binding to a PDZ domain ligand from another protein [13, 14].

Because of the specificity and high affinity of PDZ domains for their C-terminal ligands and for PDZ domains containing β-hairpin ligands, we reasoned that these interacting pairs could provide the core components of an affinity chromatography system, including affinity resins, affinity tags, and elution agents. Previously, the PDZ1 domain of the *Drosophila* InaD protein was linked to a support resin and used to purify proteins engineered to contain the NorpA PDZ1 peptide ligand [15]. This ligand includes a cysteine residue that attaches covalently via a disulfide bond to the InaD PDZ1 binding pocket. Proteins were eluted from the InaD affinity column by exposure to a reducing reagent. Here we show that affinity columns containing the PDZ domains of PSD-95 can be used to purify active PDZ domain-binding proteins to very high purity in a single step without disulfide bond formation. We prepared solid supports derivatized with recombinant PDZ domains from PSD-95 and used them to purify five heterologously expressed neuronal proteins that contain endogenous PDZ domain ligands (Table 2.1), eluting them with synthetic peptides that have the sequences of cognate PDZ domain ligands.

We also show that addition of PDZ domain-related affinity tags to POIs that do not contain endogenous PDZ domains or ligands enables purification of the POIs on the affinity resins. We used peptides derived from the C-terminal PDZ domain ligand of the N-methyl-D-aspartate type glutamate receptor (NMDAR) subunit GluN2B, PDZ domains from PSD-95, and the nNOS PDZ β-Hairpin domain (PDZbh) to construct affinity tags

and corresponding affinity resins. We verified that the tags do not alter protein activity by fusing PDZ domain-derived affinity tags to DasherGFP, β-Galactosidase (LacZ), and chloramphenicol acetyltransferase (CAT) and then assaying their activity using standardized spectrophotometric and fluorescence assays.

MATERIALS AND METHODS

Bacterial Strains and Materials

TOP10 cells (Cat. No. C4040-10, Life Technologies, Carlsbad CA) were used for plasmid DNA propagation and cloning, and BL21(DE3) cells (Cat. No. 69450-4l, EMD Millipore, Billerica MA) for protein expression. PfuUltra II Fusion HS DNA Polymerase (Cat. No. 600670, Agilent, Santa Clara CA) was used for Polymerase Incomplete Primer Extension-Ligation Independent Cloning (PIPE-LIC) and for Polymerase Incomplete Primer Extension Mutagenesis (PIPE-Mutagenesis) [16, 17]. All restriction enzymes (PvuI, Cat. No. R3150S; SbfI, Cat. No. R3642S) and DNA ligase (QuickLigase, Cat. No. M2200S) were purchased from NEB (Ipswitch MA). Chloramphenicol (Cat. No. C0378), acetyl-CoA (Cat. No. A2056), 5,5'-dithiobis-2-nitrobenzoic acid (DTNB) (Cat. No. D8130), CAT enzyme (Cat. No. C8413), *o*-nitrophenol-β-galactoside (ONPG) (Cat. No. N1127), *o*-nitrophenol (ON) (Cat. No. N19702), reducing agents tris (2-carboxyethyl) phosphine hydrochloride (TCEP) (Cat. No. C4706) and β-mercaptoethanol (Cat. No. M6250), and buffers were purchased from Sigma Aldrich (St. Louis MO). Peptides used for elution of bound protein from affinity resin

(YKQTSV and SIESDV) or for coupling to affinity resin (GAGSSIESDV) were purchased from Genscript (Piscataway NJ).

Expression Vectors and Cloning

The following cDNAs were codon-optimized for expression in *E. coli*, synthesized, and inserted into the pJExpress414 expression vector (Cat. No. pJ414) by DNA 2.0 (Menlo Park–CA): *Mus musculus* GluN2B (Residues 842 to 1482; **Q01097**) fused to maltose binding protein (MBP-GluN2B Tail); *E. coli* thioredoxin (THX) (**P0AA25**); Dasher Green Fluorescent Protein (DasherGFP; Custom Order, DNA2.0, Menlo Park CA); *Rattus norvegicus* CRIPT (**Q792Q4**); residues 11 to 129 of *Mus musculus* nNOS (**Q9Z0J4**), containing the PDZ domain (single PDZbh) or two fused copies of the PDZ domain (tandem PDZbh); and PDZ1 (residues 61 to 151), PDZ2 (residues 155 to 249), PDZ3 (residues 302 to 402) and PDZ1-2 (residues 61 to 249) of *Mus musculus* PSD-95 (**Q62108**). A cDNA encoding MBP in pMAL-c4e plasmid (Cat. No. 8110) was purchased from New England Biolabs (Ipswitch MA) and a cDNA encoding glutathione S-transferase (GST) in pGEX-6P-1 plasmid (Cat. No. 28-9546-48) was purchased from General Electric Healthcare (Uppsala, Sweden). The two were transferred into pJExpress414 plasmids containing protein domains via PIPE-LIC to generate expression vectors encoding the proteins fused to MBP, GST, or MBP-GST tags. A cDNA encoding *E. coli K12* β-Galactosidase (LacZ) (**P00722**) in pMS34 plasmid (Cat. No. 32297) was purchased from AddGene (Cambridge MA). A cDNA encoding CAT (**R0JP21**/**Q5YFS3**/**Q4VY50**) in pCAT3-Basic plasmid (Cat. No. E1871) was purchased

from Promega (Madison WI). cDNAs encoding *Rattus norvegicus* synGAP (**Q9QUH6**) and *Homo sapiens* cypin (**Q9Y2T3**) in pCMV6 plasmids (Cat. No. RN200179 and RC208771, respectively) were purchased from OriGene (Rockville MD) and were transferred into pJExpress414 expression vectors containing an N-terminal 6xHis Tag (synGAP residues 103 to 1293), MBP Tag (cypin), or no tag (cypin) via PIPE-LIC.

To generate expression vectors for the production of PDZ domain-HaloTag fusion proteins, genes encoding the PDZ domains from PSD-95 (PDZ1, PDZ2, PDZ3, PDZ1-2) and nNOS (PDZbh) were excised from pJExpress414 with PvuI and SbfI restriction enzymes, separated by agarose electrophoresis, gel-purified, and ligated into gel-purified pFN18A (Cat. No. G2751, Promega, Madison WI) digested with PvuI and SbfI.

To generate affinity tags derived from C-terminal peptide PDZ domain ligands, cDNAs encoding fusion proteins with an N-terminal GST, MBP, or THX were cloned into pJExpress414 in frame to DNA encoding the C-terminal ten amino acids of GluN2B (EKLSSIESDV), termed GluN2B ligand, by PIPE-LIC. Specific truncations of zero to ten amino acids of the GluN2B ligand fused in frame to an N-terminal GST, MBP, or THX were generated by PIPE-Mutagenesis.

To generate internal, N-terminal, or C-terminal PDZ domain or PDZbh domain affinity tags, cDNAs encoding the PDZ2 domain from PSD-95, or single or tandem PDZbh domains from nNOS, were cloned into pJExpress414 plasmids in-frame with cDNAs encoding MBP, GST, or MBP-GST fusion proteins via PIPE-LIC. PDZ domain affinity tags were inserted N-terminally or C-terminally into MBP, or internally into the MBP-GST fusion protein.

cDNAs encoding DasherGFP (in pJExpress414), CAT (in pCAT3-Basic) and LacZ (in pMS34) were fused in frame with cDNA encoding the GluN2B ligand affinity tag, or the N-terminal PDZ2 affinity tag, or the PDZbh affinity tag into pJExpress414 plasmids by PIPE-LIC cloning.

Expression of Recombinant Proteins

Single colonies of BL21(DE3) cells harboring a pJExpress414 or pFN18A plasmid were grown overnight at 37 °C in 5 ml of lysogeny broth (LB) (Cat. No. L9110, Teknova, Hollister CA) supplemented with 100 μg/ml carbenicillin. Unless specified otherwise, overnight cultures were diluted 1:500 into Overnight Express Instant Terrific Broth (TB) Media (Cat. No. 71491, EMD Millipore, Billerica MA), grown for 16 hours at 37 °C, pelleted by centrifugation, and flash frozen in liquid nitrogen.

Cells expressing synGAP or MBP-GluN2B ligand were grown in LB medium at 37 °C until cultures reached an $O.D._{600}$ of 0.8. Cultures were then chilled to 18°C, and IPTG was added to a final concentration of 0.2 or 1 mM, for synGAP or the MBP-GluN2B ligand, respectively. Cultures were grown for an additional 24 hours at 18 °C.

Synthesis of PDZ Domain-HaloTag-HaloLink Affinity Resin

We describe the synthesis of 10 ml of PDZ Domain-HaloTag-HaloLink resin. This procedure was followed for synthesis of affinity resins linked to PDZ1, PDZ2, PDZ1 plus PDZ2, and PDZ3 from PSD-95, and for affinity resins linked to single and tandem PDZbh. Bacterial cell pellets containing PDZ domain-HaloTag fusion proteins were

resuspended in 10 ml of Purification Buffer [50 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM TCEP, 0.5 mM EDTA, 10 mM $MgCl_2$, 2 mM phenylmethylsulfonyl fluoride (PMSF) (Cat. No. 78830, Sigma Aldrich, St. Louis MO), Complete Protease Inhibitor (Cat. No. 04693116001, Roche, Indianapolis IN)], supplemented with 25 U/ml Benzonase (Cat. No. 71206, EMD Millipore, Billerica MA) and 10 U/ml ReadyLyse (Cat. No. R1810M, Epicentre, Madison WI), per gram of cells. Cells were evenly suspended in a Teflon-glass homogenizer and then lysed by three passes through a ML-110 microfluidizer (Microfluidics, Westwood MA). The cell lysate was clarified by centrifugation at 30,000 × g for 60 minutes at 4°C. The clarified lysate was added to 10 ml of settled HaloLink resin (Cat. No. G1915, Promega, Madison WI) that had been pre-equilibrated in Purification Buffer and then mixed with continuous agitation for 1.5 hours at 4°C on an end-over-end mixer. Unbound protein was separated from the derivatized resin by centrifugation at 2,000 × g for 5 minutes at 4°C. The PDZ-HaloTag-HaloLink resin was then resuspended in 1 column volume of Purification Buffer and transferred to a Glass Econo-Column Chromatography Column (Cat. No. 737-0717, BioRad, Hercules CA), capped, and allowed to settle for 60 minutes before the cap was removed. The resin was washed with 20 column volumes of Purification Buffer and then with 20 column volumes of Purification Buffer supplemented with 0.05% NaN_3. It was resuspended in 1 column volume of Purification Buffer supplemented with 0.05% NaN_3, and stored at 4°C.

To determine the density of PDZ domains on the resin, we released the PDZ domains from the resin by digestion of 200 µl of settled resin with ProTEV Plus (Cat. No. V6102, Promega, Madison WI). We determined the amount of protein released by SDS-

PAGE as described below under "Assessment of Protein Purity and Yield." The densities varied from 45 to 350 pmol PDZ domain/µl resin (Table 2.2).

Synthesis of GluN2B ligand-NHS-Agarose Affinity Resin

We describe the synthesis of 10 ml of GluN2B ligand-NHS-Agarose. Twenty ml of a 50% slurry of N-hydroxysuccinimide (NHS)-Activated Agarose Resin in acetone storage solution (Cat. No. 26200, Pierce, Rockford IL) was added to a 50 ml Falcon Tube. The storage solution was removed by centrifugation at 1,000 × g for 1 minute. The 10 ml of settled NHS-Activated Agarose resin was washed twice with 3 column volumes of ultrapure water followed by centrifugation at 1,000 × g for 1 minute. The NHS-Activated Agarose was then washed twice with 3 column volumes of Coupling/Wash Buffer (100 mM $NaHPO_4$, pH 7.2, 150 mM NaCl) followed by centrifugation at 1,000 × g for 1 minute. Peptide Buffer (2.5 column volumes of 100 mM $NaHPO_4$, pH 7.2, 150 mM NaCl, 10 mg/ml GAGSSIESDV peptide) was added to the resin, and the mixture was incubated for 2 hours at room temperature with continuous agitation on an end-over-end mixer. The GluN2B ligand-NHS-Activated Agarose was pelleted by centrifugation at 1,000 × g for 1 minute, and washed twice with 3 column volumes of Coupling/Wash Buffer followed by centrifugation at 1,000 × g for 1 minute. The remaining active sites on the GluN2B ligand-NHS-Agarose were blocked by incubation with 3 column volumes of Quenching Buffer (1 M ethanolamine, pH 7.4) for 20 minutes at room temperature with end-over-end mixing. Quenching Buffer was removed by centrifugation at 1,000 × g for 1 minute. GluN2B ligand-NHS-Agarose resin was then washed twice with 4 column volumes of Coupling/Wash Buffer separated by centrifugation at 1,000 × g for 1 minute. For storage,

the GluN2B ligand-NHS-Agarose was washed twice with 4 column volumes of Storage Buffer (100 mM $NaHPO_4$, pH 7.2, 150 mM NaCl, 0.05% NaN_3), pelleted by centrifugation at 1,000 × g for 1 minute, resuspended in 1 column volume of Storage Buffer, and stored at 4°C.

Coupling of peptide to NHS-Agarose and the ligand density of immobilized peptide were monitored by 660 nm Protein Assay (Cat. No. 22660, Pierce, Rockford IL) (The NHS leaving group interferes with BCA assays and absorbance assays at 280 nm). Ligand densities varied from 20 to 26 nmol peptide/µl resin (20-26 mM).

Purification of PDZ2 Domain of PSD-95 and Synthesis of PDZ Domain-NHS-Agarose Affinity Resin

We describe the purification of the PDZ2 domain of PSD-95 and the synthesis of 5 ml of PDZ domain-NHS-Agarose resin. The PDZ2 domain of PSD-95 was expressed in *E. coli* as described above and purified by chromatography on GluN2B ligand-NHS-Agarose Affinity Resin (synthesis described above). Bacterial cell pellets containing recombinant PDZ2 domains were resuspended in 10 ml of Purification Buffer, supplemented with 25 U/ml Benzonase and 10 U/ml ReadyLyse, per gram of cells. Cells were evenly suspended in a Teflon-glass homogenizer and then lysed by three passes through a Microfluidizer. The cell lysate was clarified by centrifugation at 30,000 × g for 60 minutes at 4°C. The clarified lysate was added to 15 ml of settled GluN2B ligand-NHS-Agarose that had been pre-equilibrated in Purification Buffer, and the suspension was incubated for 1 hour at 4°C with continuous agitation on an end-over-end mixer.

Unbound protein was separated from the affinity resin by centrifugation at 2,000 × g for 2 minutes at 4°C. The resin was resuspended in 1 column volume of Purification Buffer and transferred to a Glass Econo-Column Chromatography Column, capped, and allowed to settle for 60 minutes before the cap was removed. The resin was washed with 20 column volumes of Purification Buffer. The PDZ2 domain protein was eluted by application of 4 column volumes of Peptide Elution Buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM TCEP, Complete Protease Inhibitor, and 400 μg/ml SIESDV peptide). Aliquots (17.5 ml) of eluted PDZ2 domain protein were pipetted into 20 ml Pierce 9K Protein Concentrators (Cat. No. 89885A, Thermo Scientific, Rockford IL), which were then subjected to centrifugation at 3,000 × g for 25 minutes at 4°C in a swinging bucket rotor. The filtrate was discarded and another 17.5 ml aliquot of eluted PDZ2 domain protein was added to the concentrators and subjected to centrifugation. These steps were repeated until the final total volume of PDZ2 domain protein eluate was reduced to 10 ml. The concentrated PDZ2 domain pool (11.3 mg/ml, 723 μM) was subjected to 5 cycles of buffer exchange into Coupling/Wash Buffer in preparation for coupling to NHS-Activated Agarose. The buffer exchange also served to remove the bound GluN2B ligand elution peptide from the purified PDZ2 domains.

Purified PDZ2 domains in Coupling/Wash Buffer (11.3 mg/ml, 723 μM) were coupled to NHS-Activated Agarose resin by the same procedure described for coupling of the GluN2B ligand to NHS-Activated Agarose resin. The density of PDZ2 domains on the resin was approximately 0.724 nmol protein/μl resin (724 μM; 9.4 mg/ml).

Protein Purification on PDZ2, PDZ1 plus PDZ2, and PDZ3-HaloTag-HaloLink Resin, Single and Tandem PDZbh Resin, PDZ2-NHS-Agarose, and GluN2B ligand-NHS Agarose Resin

Unless otherwise specified, all bacterial cell pellets containing tagged or untagged recombinant POIs were resuspended in 5 ml/g cell paste of BugBuster Lysis Buffer [1x BugBuster (Cat. No. 70921, EMD Millipore, Billerica MA), 50 mM HEPES, 150 mM NaCl, 5 mM TCEP, 0.5 mM EDTA, 10 mM $MgCl_2$, 2 mM PMSF, Complete Protease Inhibitor supplemented with 25 U/ml Benzonase and 10 U/ml ReadyLyse]. Resuspended cells were incubated on a shaking platform at low speed for 20 minutes at room temperature before the lysate was clarified by centrifugation at 16,000 × g for 30 minutes at 4°C.

All protein purification was performed by gravity-flow, batch chromatography. Clarified lysate (1.8 ml) containing the POI was added to 200 µl of resin pre-equilibrated with Purification Buffer, and the suspension was agitated at 4 °C for 1 hour on an end-over-end mixer to allow the protein to bind to the affinity resin. The mixtures were then subjected to centrifugation at 1,000 × g for 2 minutes, and the supernatant was removed by pipetting. The affinity resin with bound protein was then resuspended in 2 column volumes of Purification Buffer, transferred to a BioSpin Chromatography Column (Cat. No. 732-6008, BioRad, Hercules CA) capped, and allowed to settle for 15 minutes. The cap was removed and the resin washed with 20 column volumes of Purification Buffer. Bound protein was eluted by application of 4 column volumes of Peptide Elution Buffer

[50 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM TCEP, Roche Complete Protease Inhibitor, and 400 μg/ml YKQTSV peptide (for PDZ3 Domain-resin) or SIESDV peptide (for PDZ2 Domain-resin, PDZ1 plus PDZ2 Domain-resin, and for single and tandem PDZbh Domain-resin)]. The concentration of the eluted protein could be increased by decreasing the volume of buffer added for elution, increasing the concentration of peptide from 400 to 800 μg/ml, and/or by incubating the column resin with Peptide Elution Buffer without flow for 20 minutes before collecting the eluate.

Remaining bound protein (<10% of bound protein before elution) and protein contaminants were removed from the resin by applying 4 column volumes of Denaturing Buffer (8 M Urea). The resin was then washed with 15 column volumes of Storage Buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM TCEP, 0.05% NaN_3) to renature the immobilized PDZ domains or GluN2B ligand. Regenerated resin was stored in BD Falcon Tubes (Cat. No. 352070, BD Biosciences, Bedford MA) at 4°C. No loss of binding capacity was observed after 5 cycles of elution, denaturation and renaturation, or after storage for 6 months at 4°C.

Bacterial cell pellets containing the MBP-GluN2B ligand fusion protein, synGAP, and fusions of PDZ domains to functional proteins (DasherGFP, LacZ, and CAT) were resuspended in 5 ml/g cell-paste Purification Buffer (50 mM HEPES, pH 7.5, 150 mM NaCl, 5 mM TCEP, 0.5 mM EDTA, 10 mM $MgCl_2$, 2 mM PMSF, Complete Protease Inhibitor) supplemented with 25 U/ml Benzonase and 10 U/ml ReadyLyse. The resuspended cells were sonicated (Branson, Danbury CT) 2x for 90 seconds/pass (15% power, 1.0 second on, 1.5 seconds off) before clarification by centrifugation at 16,000 × g

for 30 minutes at 4°C. In addition, all buffers that we used for purification of synGAP contained 0.02% Tergitol Type NP-40 (Cat. No. NP40S, Sigma-Aldrich, St. Louis MO). Otherwise, the purification methods were as described for the other POIs.

Assessment of Protein Purity and Yield

We used SDS-PAGE to determine purity of proteins and to quantify yields throughout the purification process. Protein samples were diluted into 4x LDS buffer (Cat. No. NP0008, Life Technologies, Carlsbad CA) and heated to 70°C for 10 minutes before fractionation on NuPAGE Novex 4-12% Bis-Tris SDS-PAGE gradient gels (Cat. No. NP0008, Life Technologies, Carlsbad CA) run under reducing conditions with NuPAGE Antioxidant (Cat. No. NP0005), and MOPS or MES running buffers (Cat. Nos. NP0001 and NP0002), all purchased from Life Technologies (Carlsbad CA). Proteins in the gels were stained with Gel Code Blue (Cat. No. 24592, Thermo Scientific, Rockford IL), and imaged on a Licor Odyssey Imager (Licor Biosciences, Lincoln NE) at 700 nm. Molecular weights of stained proteins were verified by comparison to Precision Plus Protein All Blue Standards (Cat. No. 161-0373, BioRad, Hercules CA). The amount of each stained protein was measured in the gel by comparison to known amounts of bovine serum albumin (Cat. No. A5611), lysozyme (Cat. No. L6876), or LacZ (Cat. No. G8511), all purchased from Sigma-Aldrich (St. Louis MO). The protein standards were loaded onto each gel in lanes adjacent to the protein samples.

Assays of Protein Function

The integrity of the purified DasherGFP fusion protein's fluorophore was assayed on a Hitachi F-4500 FL Fluorescence Spectrometer (Tokyo, Japan) at 22°C. Excitation (300-600 nm) and emission (350-700 nm) spectra were collected to verify fluorophore activity. The fraction of properly folded DasherGFP fusion proteins was measured by comparing the emitted fluorescence intensity at 520 nm of tagged, purified DasherGFP fusion proteins excited at 505 nm to that of an equimolar amount of untagged, purified DasherGFP.

The activity of purified LacZ fusion proteins was assayed by measuring the rate of hydrolysis of ONPG on a VERSAmax tunable microplate reader at 20°C as described in [18]. Briefly, the rate of ON release after addition of LacZ, produced by hydrolysis of ONPG, was measured by continuous monitoring of absorbance at 420 nm. Initial rates of ONPG hydrolysis were entered into Prism (v6.0d, GraphPad Software, La Jolla CA), plotted against ONPG substrate concentrations and analyzed by nonlinear regression using the Michaelis-Menten equation to calculate k_{cat} and K_M values. The activity of purified LacZ fusion proteins was compared to literature values [19]. The purified LacZ fusion proteins were prepared for assay by buffer exchange in an Amicon Ultra-0.5 ml concentrator into 50 mM Tris, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 0.01 mM $MnCl_2$, and 100 mM β-mercaptoethanol [20]. Aliquots were flash frozen in liquid nitrogen and stored at -80°C.

The activity of purified CAT fusion proteins was determined by measuring the rate of production of 5-thio-2-nitrobenzoic acid resulting from the reduction of DTNB by free CoA generated during acylation of chloramphenicol. The rate of reduction was measured by continuously monitoring absorbance at 412 nm after addition of CAT on a VERSAmax tunable microplate reader at 37 °C using a DTNB reduction assay described in [21]. Initial rates of chloramphenicol acylation were entered into Prism (v6.0d; GraphPad Software, La Jolla CA), plotted against acetyl-CoA or chloramphenicol substrate concentrations and analyzed by nonlinear regression, using the Michaelis-Menten equation to calculate k_{cat} and K_M values. The activity of purified CAT fusion proteins and equimolar amounts of untagged CAT (Sigma) were compared. CAT fusion proteins were prepared for assay by buffer exchange in an Amicon Ultra-0.5 ml concentrator into10 mM Tris, pH 7.8, 200 mM NaCl, 0.2 mM chloramphenicol, 0.1 mM β-mercaptoethanol [22]. Aliquots were flash frozen with liquid nitrogen and stored at -80°C.

RESULTS/DISCUSSION

Synthesis and Testing of Affinity Resins

Synthesis of PDZ Domain and PDZ Domain Binding Peptide Affinity Resin

Figure 2.1 outlines one possible workflow for PDZ Affinity Chromatography. A protein of interest (POI) containing a C-terminal PDZ domain peptide ligand is expressed in *E. coli*. The ligand can be endogenous or introduced by genetic engineering. Clarified bacterial lysate containing the POI is passed over a solid support resin derivatized with a recombinant PDZ domain. The PDZ domain ligand on the POI directs binding to the PDZ

domain immobilized on the resin. After washing to remove loosely bound contaminants remaining from the cell lysate, free peptide ligand is added to the resin to competitively remove the bound POI from the immobilized PDZ domains, resulting in elution of a purified POI. In an alternative scheme, a POI containing a PDZ domain is passed over a solid support derivatized with a PDZ domain peptide ligand and eluted by addition of the free ligand (Fig. 2.2). In both of these schemes, recombinant affinity tags can be introduced with a TEV protease site between the tag and the POI, permitting removal of the tag by proteolysis to generate purified untagged protein.

To prepare affinity resin containing immobilized PDZ domains, we chose to use the HaloTag and HaloLink resin manufactured by Promega because of its speed and ease of use compared to the conventional NHS coupling method, and also because it permits coupling with a single, uniform orientation. The HaloTag is a catalytically inactive bacterial haloalkane dehalogenase from *Rhodococcus* that has been engineered to rapidly form a covalent attachment to synthetic chloroalkane ligands [23, 24]. The HaloTag and HaloLink resin have been used for rapid, small scale preparation of affinity resins to purify polyclonal antibodies [25]. Using a similar procedure (described in Methods), we expressed HaloTag-PDZ domain fusion proteins in *E. coli*, lysed the cells, and incubated the clarified lysate with HaloLink resin to generate the PDZ domain-HaloTag-HaloLink affinity resin (Fig. 2.3). Coupling of PDZ Domain-HaloTag fusion proteins to HaloLink resin was monitored on SDS gels by determining the amount of PDZ domain protein released from the resin after cleavage of the HaloLink with TEV protease, as described in Methods. Coupling densities varied from 45 to 350 pmol PDZ domain/μl resin (Table 2.2).

In addition to preparing PDZ Domain Affinity Resin, we also prepared resin derivatized with PDZ domain peptide ligands for use in purifying POIs containing PDZ domains. Peptides were coupled to Agarose with an NHS linker, as described in Methods. The densities of peptide ligands on NHS coupled resin varied from 20 to 26 nmol peptide/µl resin.

Purification of the Cytosolic Domain of GluN2B

To test the utility of PDZ Domain Affinity Resin, we purified a panel of proteins known to interact with the PDZ domains of PSD-95 (Table 2.1). PSD-95 is a scaffold protein located in the postsynaptic density of glutametergic synapses that contains three PDZ domains in addition to an SH3 and a guanylate kinase-like domain [1, 4]. PSD-95 is the first protein in which PDZ domains were recognized. Because of its critical role in receptor clustering and protein localization at the synapse [26-29], it is also one the most well characterized PDZ domain-containing proteins.

The N-methyl-D-aspartate type glutamate receptor (NMDAR) is one of three major classes of receptors for glutamate, the principal excitatory neurotransmitter in the central nervous system and has a seminal role in learning and memory through its actions as a "coincidence detector" that initiates changes in synaptic strength leading to the formation of new neural networks [30-33]. It is comprised of four subunits, including two from the GluN2 subfamily (GluN2A-D) [34]. The GluN2 subunits are unique among ligand-gated channels in that they have extended intracellular C-terminal domains or "tails," which extend into the cytoplasm and serve as nucleation sites for scaffolds and signal transducing

enzymes [35]. The cytoplasmic tails of NR2 subunits have not previously been expressed in soluble form in significant quantities, or purified. We chose to express and purify the cytoplasmic tail of the GluN2B subunit ("GluN2B Tail": GluN2B residues 840-1482) which contains a C-terminal PDZ ligand that binds to PDZ2 of PSD-95. A PDZ Domain Affinity Resin substituted with PDZ2 from PSD-95 was synthesized as described in Methods. When expressed in *E. coli*, the GluN2B Tail forms insoluble protein aggregates known as inclusion bodies (Supp. Fig. 2.1A); however, we found that it could be expressed in soluble form when fused at its N-terminus to MBP. The bacterial lysate containing the MBP-GluN2B Tail fusion protein was incubated with PDZ2 Domain Affinity Resin, and eluted with a competing peptide ligand as described in Methods. The eluted fusion protein was >70% pure (Fig. 2.4A) and was recovered at a concentration of 0.5 μM with a 6% yield. We found that PDZ Domain Affinity Chromatography was the method of choice for purifying the GluN2B Tail, compared to alternative methods. For example, after a similar column protocol substituting Amylose (MBP Tag) Affinity resin, the MBP-GluN2B Tail fusion protein was only approximately 8% pure, and was recovered at a concentration of 0.1 μM with a yield of 6% (Supp. Fig. 2.1B). Attempts to purify the MBP-GluN2B Tail fusion protein with Streptactin and Ni-NTA resins after addition of N- or C-terminal StrepII or 8xHis Tags, respectively, were unsuccessful; no protein was present in the column eluate (data not shown).

Purification of Soluble Recombinant SynGAP

SynGAP (Synaptic GTPase Activating Protein) is a dual Ras and Rap GTPase

activating protein that is highly concentrated in the postsynaptic density of excitatory synapses [36]. Homozygous deletion of synGAP is lethal in mice [37, 38], and a heterozygous deletion confers behavioral phenotypes associated with cognitive disability and mental illness [39]. Synaptic plasticity is disrupted in heterozygous knockout mice and the formation of spine synapses during development is accelerated [38-40]. Mutations in the human SYNGAP1 gene appear to cause non-syndromic intellectual disability and certain forms of autism [41]. SynGAP is regulated by phosphorylation in a disordered region carboxyl to the GAP domain [42]. Attempts to understand the effects of phosphorylation have been hampered by the difficulty of purifying a soluble form of synGAP containing the disordered region. It was previously shown that removal of the N-terminal 102 residues of synGAP prevents the association with membranes of a recombinant form containing the PH, C2, and GAP domains (residues 103-725), but lacking the disordered region [43]. Therefore, we attempted to express and purify a soluble form of synGAP that also contains the disordered region beyond the GAP domain (r-synGAP; residues 103-1293). We found that r-synGAP remains soluble after expression in *E. coli*. This form of synGAP contains a carboxyl terminal ligand that binds to PDZ3 of PSD-95. Therefore, we synthesized PDZ3 Domain Affinity Resin and used it to purify r-synGAP to >70% homogeneity with a concentration of 0.2 μM and a yield of 7% (Fig. 2.4B), as described in Methods. When r-synGAP was tagged with an N-terminal 6x Histidine tag and purified by a similar procedure on Talon Affinity Resin, the purity of the recovered protein was less, approximately 25%, but the yield was considerably greater (75% at a concentration of 2 μM) than after PDZ3 Domain Affinity Chromatography

(Supp. Fig. 2.1C). In addition to an appended N-terminal 6x Histidine tag, r-synGAP contains a naturally occuring internal 10x polyhistidine domain which markedly increases its affinity for Talon resin; we believe this contributes to its high yield from the Talon resin.

Purification of CRIPT (Cysteine Rich Interactor of PDZ Three)

CRIPT (Cysteine Rich Interactor of PDZ Three) is a small (101 residue), highly conserved, cysteine-rich protein concentrated in the PSD of glutamatergic synapses in pyramidal neurons [9, 44, 45]. CRIPT is one of three proteins known to bind to PDZ3 of PSD-95. It also binds the homologous PDZ3 domains of PSD-93 (chapsyn-110), SAP102 and SAP97 [9]. It can bind simultaneously to polymerized tubulin, thus linking PSD-95 family proteins to the microtubule cytoskeleton [44]. CRIPT contains 8 cysteine residues, all confined to CXXC repeats. To test whether PDZ3 Affinity Chromatography can be carried out in the presence of reducing agents, we chose to use it to purify recombinant CRIPT. CRIPT was previously purifed as a THX- fusion protein [44], so we attempted to purify it in the absence of exogenous affinity tags. We carried out the purification as described in Methods, except that, in one experiment, a reducing agent (5 mM TCEP) was present in all buffers; in a second the TCEP was not included. The results of the two purifications were identical. In the presence or absence of TCEP, CRIPT was purified to >95% homogeneity, with a yield of 35% and a concentration of 9 μM (Fig. 2.4C). Thus, unlike the NorpA-InaD affinity chromatography system [15], affinity chromatography with the PDZ domains from PSD-95 can be carried out either in the presence or absence of

reducing agents.

Purification of PDZbh from nNOS (Neuronal Nitric Oxide Synthetase)

Neuronal nitric oxide synthase (nNOS) [46, 47] catalyzes the formation of nitric oxide and citrulline from arginine. Activation of nNOS by binding of Ca^{2+}/calmodulin leads to formatiom of NO, which in turn activates guanylyl cyclase to form cGMP. NO generation can modulate synaptic plasticity by acting as a retrograde synaptic transmitter, and, in some circumstances, provide neuroprotection against excitoxicity [48]. It contains a single PDZ domain (Residues 14-125 of *Mus musculus* nNOS) with an unusual structure that includes a β-Hairpin (bh) fold that itself acts as a PDZ domain ligand [13]. We were able to express the PDZbh domain from nNOS in *E. coli* and purify it to >95% homogeneity on the PDZ2 Domain Affinity Resin with a yield of 1 percent, as described in Methods (Fig. 2.4D) (in retrospect, we realized that the low yield was caused by addition of a 40-fold molar excess of expressed PDZbh to the resin.) The concentration of purified PDZbh was 7.6 μM.

Purification of Cypin

Cypin was originally discovered and characterized as an abundant cytosolic protein that interacts directly with PSD-95 [49]. When overexpressed in neurons, it appears to trap PSD-95 in the cytosol and reduces the targeting of PSD-95 to synapses. During neural development, it contributes to regulation of dendritic branching, in part by interfering with the interaction between PSD-95 and microtubules [50]. Cypin binds to the

PSD-95 family of proteins via a C-terminal SSSV sequence and has the unique property that it requires the presence of both PDZ1 and PDZ2 for detectable binding [49]. Cypin has been partially purified from brain extracts in small quantities by affinity purification with glutathione-beads bound to GST-PSD-95 [49]. More recently, both GST- and MBP-fusion proteins of cypin have been purified by standard affinity chromatography [51].

We used an affinity column substituted with a HaloTag-PDZ1 plus PDZ2 fusion protein (residues 61 to 249 of *Mus musculus* PSD-95) to purify heterologously expressed human cypin, as described in Methods. It was expressed in *E. coli* as the free protein and also fused to an N-terminal MBP, and purified from the bacterial supernate on the PDZ1-PDZ2 Domain-Affinity Resin, as described in Methods. Both cypin and MBP-cypin were purified to >99% homogeneity (Fig. 2.4E) with yields of 95% and 70% and concentrations of 0.8 μM and 0.5 μM, respectively.

In summary, we purified five individual proteins using three different synthetic PDZ Domain Resins derivitized with PDZ domains from PSD-95. MBP-GluN2B Tail and PDZbh of nNOS were purified on a PDZ2 Domain Resin. R-synGAP and CRIPT were purified on PDZ3 Domain Resin. Finally, cypin and MBP-cypin were purified on a tandem PDZ1-PDZ2 Domain Resin. We successfully completed all of the purifications that we attempted. CRIPT, which contains several labile cysteines, was purified in the presence of TCEP, demonstrating that the PDZ Domain Resins and their respective ligands are compatible with the use of reducing agents. Our results demonstrate that PDZ Domain Affinity Chromatography is an excellent tool for the rapid purification of proteins with endogenous PDZ domain ligands.

Comparison of NHS and the HaloTag-HaloLink System for synthesis of PDZ Domain-Affinity Resin

We used the HaloTag-HaloLink system to synthesize most of the PDZ Domain Affinity Resins discussed here. However, we recognize that the cost of preparing HaloLink-HaloTag-PDZ domain resin would be prohibitive for large-scale applications in which yields of tens to hundreds milligrams of protein are required. Therefore, we tested a less expensive linkage method for preparation of the PDZ Domain Affinity Resins. We purified the recombinant PDZ2 domain from PSD-95 on a PDZ Domain Ligand Affinity Resin and coupled it to NHS-Agarose, as described in Methods (Fig. 2.5A). The purification yielded approximately 110 mg of 40 μM PDZ2 domain protein. The purified PDZ2 domain was concentrated to 723 μM in preparation for coupling to NHS-Agarose, as described in Methods. Residual PDZ Domain Peptide Ligand was then removed by five exchanges of buffer by ultrafiltration, each time diluting the concentrated protein with buffer that did not contain the peptide. The concentrated PDZ2 domain, freed of ligand, was coupled to NHS-Agarose, and unreacted NHS sites were quenched with ethanolamine as described in Methods. Coupling was monitored on SDS gels, as described in Methods. The ligand density of PDZ2 domains on the coupled Agarose was 724 pmol/μl settled resin (724 μM; 9.4 mg/ml).

We compared PDZ2 Domain Affinity Resin coupled by NHS to that coupled by the HaloTag-HaloLink method by purifying a test panel of three proteins that we also purified using HaloLink-HaloTag-PDZ2 Domain Affinity Resin: MBP-GluN2B Tail, and MBP

and GST both fused to a PDZ2 Ligand Affinity Tag as described below and in Methods (Fig. 2.5B and C). All three proteins were purified to >95% homogeneity (Fig. 2.5C). Interestingly, we obtained higher yields for each of the proteins when the PDZ2 Domain was linked to Agarose by NHS: MBP-GluN2B Tail (1.2 μM vs. 0.5 μM), MBP-PDZ2 ligand (12.9 μM vs. 9.6 μM), and GST- PDZ2 ligand (22.7 μM vs. 19.5 μM). The increased yield with NHS-coupled PDZ2 Domain Resin is likely due to its higher ligand density relative to the HaloLink-HaloTag-coupled resin (724 μM vs. 266 μM, respectively). Background binding to both resins was insignificant, as assessed by incubating them with untagged MBP and GST or with bacterial proteins in the *E. coli* lysate. These results indicate that PDZ domains can be successfully coupled to NHS-Agarose resin and that such resins work well for purification of proteins that contain PDZ domain ligands.

Engineering of Recombinant Affinity Tags into POIs

PDZ Domain C-terminal Peptide Ligands as Affinity Tags

The C-terminal six residues of the GluN2B subunit of the NMDAR (SIESDV) bind with high affinity to PDZ1 and PDZ2 of PSD-95, at K_D's of 2.3 and 0.7 μM, respectively [52]. We hypothesized that fusing the C-terminal residues of GluN2B to a POI would enable it to bind to the PDZ1 or PDZ2 domains of PSD-95 immobilized on an affinity resin. To test this hypothesis, we prepared recombinant proteins with varying numbers of residues from the GluN2B Tail (EKLSSIESDV) fused to GST, MBP, or THX. To determine a minimum length of the ligand sequence required to permit purification of the

proteins on PDZ2 Domain Affinity Resin, ten fusion proteins, containing one to ten of the ligand residues, were made for each of the three proteins. We found that the five C-terminal residues (IESDV) were all required for effective binding to, and for elution from, the PDZ2 Domain Affinity Resin (Fig. 2.6B). GST fusion proteins with tags longer than five residues were purified to >95% homogeneity as assessed by SDS-PAGE (Fig. 2.6B). MBP and THX fused to the truncation library of one to ten residues were also purified to >95% homogeneity (Supp. Figs. 2.2 and 2.3). To minimize the possibility of structural interference in fusion proteins containing the PDZ2 Ligand Tag, we decided to use the seven residues from the C-terminus of GluN2B, SSIESDV, as a standard Ligand Tag.

PDZ domains as affinity tags

To assess the utility of conventional PDZ domains as affinity tags, we expressed the PDZ2 Domain of PSD-95 alone, fused N- or C-terminally to MBP, or fused internally between the MBP and GST proteins. Each fusion protein was purified with single or tandem PDZbh Domain Affinity Resins prepared with the HaloTag-HaloLink system, and with PDZ2 Domain Ligand Affinity Resin prepared with NHS-Agarose, as described in Methods (Fig. 2.7A and B). All of the PDZ2 Domain Tagged POIs, regardless of the position of the tag, were purified to >95% purity on all three types of resin (Fig. 2.7B). The yields of pure protein were higher when the tags were fused C-terminally or internally than when they were fused N-terminally. Yields of tagged protein were 25- to 120-fold higher for the PDZ2 Domain Peptide Ligand Affinity Resin than for single and tandem PDZbh Domain Affinity Resins: PDZ2 domain alone (36.9 μM vs. 0.8 and 0.3 μM);

MBP-PDZ2 domain fusion (30.6 μM vs. 1.3 and 0.5 μM); and MBP-PDZ2 domain-GST fusion (19.1 μM vs. 0.8 and 0.5 μM). Control MBP or GST, expressed alone without a PDZ domain, showed no detectable binding to the Affinity Resins. This data demonstrates that PDZ Domain Peptide Ligand Affinity Resins may be particularly useful for a wide range of recombinantly tagged POIs. The utility of PDZ Domain Peptide Ligand Affinity Resins is enhanced by their low cost and ease of preparation compared to PDZbh Domain Affinity Resins.

PDZ domain beta-hairpin (PDZbh) as an affinity tag:

Because of its small, modular structure, high solubility, high expression level in *E. coli*, and its affinity for the PDZ2 Domain Affinity Resin (described under "Purification of PDZbh from nNOS"), we reasoned that the PDZbh domain from nNOS might make an excellent N-terminal, C-terminal, or internal PDZ Domain Affinity Tag. We also reasoned that two tandem PDZbh tags might have higher avidity for PDZ Domain Affinity Resins than single PDZbh tags. We tested these ideas in the experiment shown in (Fig. 2.7C and D). Positioning of a tandem PDZbh tag at the carboxyl terminus of MBP produced a slightly higher yield of the fusion protein than a single PDZbh tag when it was purified on a PDZ2 Affinity Resin. However, a single PDZbh Tag performed better than a tandem PDZbh tag when the tags were placed at the N-terminus or internally between MBP and GST in the fusion protein. PDZbh Tags placed at the C-terminus or internally produced better yields than the corresponding tags placed at the N-terminus (Fig. 2.7D).

Functional Analysis of Purified Fusion Proteins

Tagging of proteins for functional tests

CAT, LacZ, and DasherGFP proteins were recombinantly tagged with C-terminal PDZ2 Domain Peptide Ligand Affinity Tags and purified on PDZ2-NHS-Agarose resin, as described in Methods. They were purified to >95% homogeneity and recovered at concentrations of 17, 7.7, and 7.8 μM, with yields of 4, 8, and 3 % of the POI loaded on the column for CAT, LacZ, and DasherGFP, respectively (Fig. 2.8A). In these experiments, the amount of POI in the bacterial lysate loaded onto the column greatly exceeded the column capacity. We calculated that the yield of purified protein was approximately equal to the resin capacity for CAT, but about 50% of the resin capacity for LacZ and DasherGFP.

In the reverse experiment, the three proteins were tagged at the N-terminus with PDZ2 Domain Affinity Tags and purified on PDZ2 Domain Peptide Ligand Affinity Resin. They were purified to >95% homogeneity at concentrations of 70, 5.8, and 82 μM, with yields of 66, 81, and 37 % for CAT, LacZ, and DasherGFP, respectively (Fig. 2.8B). A similar amount of bacterial lysate was loaded onto each column for the purifications , as shown in Figure 2.8A and B; however, the PDZ2 Domain Ligand Resin produced a greater yield of pure protein at a higher concentration because the density of Peptide Ligand on the column was 30-fold higher than the density of the PDZ2 Domain Resin (~20 nmol/μl vs. 0.724 nmol/μl). The lower concentration of LacZ eluted from the PDZ Domain Ligand Resin and its higher yield, compared to the amount loaded, were due in part to a 13-fold reduction in the level of bacterial expression of LacZ when the PDZ Domain Ligand Affinity Tag was fused to the C-terminus rather than to the N-terminus.

These results illustrate that the optimum combination of PDZ-related Affinity Tag and Affinity Column for a particular POI will depend on the effect of the Tag on the protein's expression and functional activity, as well as on the capacity of the cognate Affinity Resin.

Finally, CAT, LacZ, and DasherGFP proteins were tagged N-terminally with single or tandem PDZbh Domain Affinity Tags and purified on PDZ2-NHS-Agarose resin. DasherGFP was purified to >95% homogeneity, but LacZ and CAT were not recovered in detectable quantities (Fig. 2.8C and D). We hypothesize that the inability to purify LacZ and CAT when fused to PDZbh Domain Affinity Tags may be caused by occlusion of the β-Hairpin domain in the fusion protein.

Function of DasherGFP

Green fluorescent protein (GFP) is a 238 residue protein isolated from the Pacific Northwest jellyfish, *Aequeorea victoria* [53]. GFP transmutes blue chemiluminescence from a primary photoprotein (aequorin) into green fluorescence [54], utilizing a *p*-hydroxybenzylidene-imidazolidone chromophore derived from its S65, Y66, and G67 residues [55]. It has been used in the production of biosensors for monitoring intracellular pH [56-58], calcium concentration [59], redox potential [60, 61], membrane potential [62], and temperature [63]. Proper folding of GFP around the chromophore is necessary for fluorescence, as evidenced by the fact that synthetic *p*-hydroxybenzylidene-imidazolidone chromophores are devoid of fluorescence [64]. When fused to the C-terminus of a POI, productive folding of the downstream GFP and formation of the fluorescent chromophore have been shown to depend on the robustness of folding of the upstream protein [65].

DasherGFP is a 26.6 kDa, synthetic, non-aequorea fluorescent protein, developed by DNA2.0, with excitation and emission wavelengths of 505 and 525 nm, respectively (Fig. 2.9A). To verify that PDZ Affinity Tags do not inhibit the fluorescence emission of DasherGFP, we performed excitation and emission wavelength scans on each purified, PDZ Affinity Tagged DasherGFP construct (Fig. 2.9). Purified DasherGFP fused to a C-terminal PDZ Domain Peptide Ligand Tag (Fig. 2.9B), an N-terminal PDZ Domain Tag (Fig. 2.9C), or a single N-terminal PDZbh Domain Tag (Fig. 2.9D) exhibited fluorescence excitation and emission spectra that are virtually identical to the unlabeled protein (Fig. 2.9A). In contrast, DasherGFP fused to an N-terminal tandem PDZbh Domain Tag (Fig. 2.9E) had greater overlap between the two spectra than the unlabeled protein, with excitation and emission maxima both occuring at 510 nm. To estimate the percentage of folded and functional tagged DasherGFP in each purified sample, the fluorescence intensity at 520 nm of each tagged DasherGFP was compared to that of an equimolar amount of untagged DasherGFP (Table 2.3). The fusions with a C-terminal PDZ Domain Peptide Ligand, an N-terminal PDZ Domain, or a single N-terminal PDZbh Domain all exhibited relative fluorescence intensities of greater than 90%, indicating that they are almost entirely folded and functional. However, the fusion with tandem PDZbh Domain Tags was almost entirely misfolded (Table 2.3), again suggesting that the tandem PDZbh Domain Affinity Tag destabilized proper folding.

Function of LacZ

β-Galactosidase (LacZ) is a 1024 residue protein isolated from *E. coli* [66, 67]. It catalyzes the cleavage of the bond between the anomeric carbon and glycosyl oxygen of a β-D-galactopyranoside [68]. *In vivo* LacZ catalyzes the cleavage of the disaccharide lactose to form glucose and galactose [69]. It is often used experimentally as a reporter of gene expression, of spontaneous or directed genetic changes in coding sequences, and of protein-protein interactions because of its activity in myriad cell lines [70-74], the availability of an array of substrates, inducers and inhibitors [74-76], its structural malleability [77-79], and the large dynamic range of its gene expression [80].

To verify that PDZ Affinity Tags do not inhibit the enzyme activity of LacZ, we assayed each tagged LacZ (Fig. 2.8) using ONPG as a substrate, as described in Methods (Fig. 2.10). The catalytic constants ($k_{cat} = V_{max}/[E]$) of LacZ fused to a C-terminal PDZ Domain Peptide Ligand (Fig. 2.10A), or an N-terminal PDZ Domain (Fig. 2.10B), were in good agreement with the published k_{cat} for untagged LacZ under identical conditions (Table 2.4). The K_M's for ONPG of these two tagged versions of LacZ were also in the same range as those published for untagged LacZ (Table 2.4). Thus, LacZ engineered to contain either of these affinity tags and purified by PDZ Affinity Chromatography is folded and functional. LacZ fused N-terminally to single or tandem PDZbh Domain Tags could not be purified by Affinity Chromatography, and so was not assayed.

Function of Chloramphenicol Acetyltransferase (CAT)

Chloramphenicol acetyltransferase (CAT) is a ~219 residue protein isolated from *E. coli* and *S. aureus*. It catalyzes the inactivation of the antibiotic chloramphenicol, by acylating chloramphenicol in the presence of acetyl-CoA to produce chloramphenicol-3-acetate and reduced CoA [81, 82]. It has been used extensively as an *in vitro* reporter of gene expression levels in eukaryotic cell lines because of its stability, absence of competing activities in eukaryotic cells, ease of use, and sensitivity [83-85].

To verify that PDZ Affinity Tags do not inhibit the enzymatic activity of CAT, we measured activity of untagged CAT and each tagged construct using chloramphenicol and acetyl-CoA as substrates, as described in Methods (Fig. 2.11). Untagged CAT (Fig. 2.11A) and CAT-fused to a C-terminal PDZ Domain Peptide Ligand (Fig. 2.11B) or an N-terminal PDZ Domain (Fig. 2.11C) were assayed with a fixed concentration of acetyl-CoA (500 μM) and 0-100 μM chloramphenicol. Their measured catalytic constants were in good agreement with the k_{cat}'s of untagged CAT purchased from Sigma Aldrich (Fig. 2.11A; Table 2.5). The K_M's for chloramphenicol of these two tagged versions of CAT are similar to those measured for untagged CAT (Table 2.5) and published values [10 μM; 21].

When the two tagged versions of CAT (Fig. 2.11E and F) were assayed with a fixed concentration of chloramphenciol (100 μM) and 0-500 μM acetyl-CoA, their catalytic constants were also in good agreement with the k_{cat}'s of untagged CAT purchased from Sigma (Fig. 2.11D; Table 2.5) and published values [50 μM; 21]. The K_M's for acetyl-CoA of these two tagged versions of CAT were also similar to published values for

untagged CAT (Table 2.5). Thus, CAT engineered to contain either of these affinity tags and purified by PDZ Domain Affinity Chromatography is folded and functional. As was the case for LacZ, CAT fused N-terminally to single or tandem PDZbh Domain Tags could not be purified by Affinity Chromatography, and so were not assayed.

CONCLUSIONS

The PDZ Domain Affinity Chromatography system reported herein adds several additional tools to the toolbox of Affinity Tags and Affinity Resins available for protein purification. When compared to the His Tag system, the C-terminal PDZ Domain Peptide Ligand Affinity Tag and PDZ Domain Affinity Resin have several advantages: the affinity of the tag for the resin is higher (~0.1 μM versus 1 μM); also, elution can be carried out with a solution of free peptide without the use of potentially denaturing elution agents, such as imidazole, and without the possibility of toxic metal ion (Zn^{2+}, Co^{2+}, Ni^{2+}) contamination by leakage off the column. A potential disadvantage of the PDZ Domain Affinity Chromatography system is that, unlike the His Tag system, it cannot be used under denaturing conditions. When compared to the GST Tag system, the PDZ Domain Affinity System also has advantages: again, the affinity of the tag for the affinity resin is higher (~0.1 μM versus 40 μM); it can be used in the presence of reducing agents; and the affinity tags are monomeric, whereas glutathione is dimeric or multimeric. A potential disadvantage of the PDZ Domain Affinity Tags relative to GST is the lack of a proven solubility-enhancing effect. Finally, the PDZ Domain-related Affinity Resins do not show detectable background binding of *E. coli* host cell proteins.

The potential to mix and match various related affinity tags and affinity resins adds flexibility to the PDZ Domain Affinity System. POIs lacking endogenous PDZ Domain Ligands can be engineered as fusion products with C-terminal PDZ Domain Peptide Ligand Affinity Tags and purified on a PDZ Domain Affinity Resin, as outlined in Fig. 2.1. Another alternative is the addition of N-terminal, C-terminal, or internal PDZ Domain Affinity Tags coupled with the use of the high capacity PDZ Domain Ligand Affinity Resin, as outlined in Fig. 2.2. The choice would depend on the expression level, stability, and solubility of the POI when fused to either of the two potential tags. For applications requiring purification of several milligrams of protein, fusion of the POI to a PDZ Domain Tag, followed by purification on a PDZ Domain Peptide Ligand-Resin, would be preferred. We have shown that CAT, LacZ, and DasherGFP fused to either an N-terminal PDZ Domain Tag or to a C-terminal PDZ Domain Peptide Ligand Affinity Tag could all be purified in fully active form on the cognate Affinity Resin.

REFERENCES

[1] H.-C. Kornau, L.T. Schenker, M.B. Kennedy, P.H. Seeburg, Domain interaction between NMDA receptor subunits and the postsynaptic density protein PSD-95. Science 269 (1995) 1737-1740.

[2] H.-C. Kornau, P.H. Seeburg, M.B. Kennedy, Interaction of ion channels and receptors with PDZ domain proteins. Curr Opin Neurobiol 7 (1997) 368-373.

[3] A.Y. Hung, M. Sheng, PDZ domains: structural modules for protein complex assembly. J Biol Chem 277 (2002) 5699-5702.

[4] K.-O. Cho, C.A. Hunt, M.B. Kennedy, The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. Neuron 9 (1992) 929-942.

[5] Z. Songyang, A.S. Fanning, C. Fu, J. Xu, S.M. Marfatia, A.H. Chishti, A. Crompton, A.C. Chan, J.M. Anderson, L.C. Cantley, Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science 275 (1997) 73-77.

[6] D.A. Doyle, A. Lee, J. Lewis, E. Kim, M. Sheng, R. MacKinnon, Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. Cell 85 (1996) 1067-1076.

[7] Y. Ivarsson, A.M. Wawrzyniak, G. Wuytens, M. Kosloff, E. Vermeiren, M. Raport, P. Zimmermann, Cooperative phosphoinositide and peptide binding by PSD-95/discs large/ZO-1 (PDZ) domain of polychaetoid, Drosophila zonulin. J Biol Chem 286 (2011) 44669-44678.

[8] J.E. Brenman, D.S. Chao, S.H. Gee, A.W. McGee, S.E. Craven, D.R. Santillano, Z. Wu, F. Huang, H. Xia, M.F. Peters, S.C. Froehner, D.S. Bredt, Interaction of nitric-oxide synthase with the postsynaptic density protein PSD-95 and a-1-syntrophin mediated by PDZ domains. Cell 84 (1996) 757-767.

[9] M. Niethammer, J.G. Valtschanoff, T.M. Kapoor, D.W. Allison, R.J. Weinberg, A.M. Craig, M. Sheng, CRIPT, a novel postsynaptic protein that binds to the third PDZ domain of PSD-95/SAP90. Neuron 20 (1998) 693-707.

[10] S.M. Marfatia, J.H. Morais Cabral, L. Lin, C. Hough, P.J. Bryant, L. Stolz, A.H. Chishti, Modular organization of the PDZ domains in the human discs-large protein suggests a mechanism for coupling PDZ domain-binding proteins to ATP and the membrane cytoskeleton. J Cell Biol 135 (1996) 753-766.

[11] B.M. Muller, U. Kistner, S. Kindler, W.J. Chung, S. Kuhlendahl, S.D. Fenster, L.F. Lau, R.W. Veh, R.L. Huganir, E.D. Gundelfinger, C.C. Garner, Sap102, a novel postsynaptic protein that interacts with NMDA receptor complexes in-vivo. Neuron 17 (1996) 255-265.

[12] B.Z. Harris, B.J. Hillier, W.A. Lim, Energetic determinants of internal motif recognition by PDZ domains. Biochemistry 40 (2001) 5921-5930.

[13] B.J. Hillier, K.S. Christopherson, K.E. Prehoda, D.S. Bredt, W.A. Lim, Unexpected modes of PDZ domain scaffolding revealed by structure of nNOS-syntrophin complex. Science 284 (1999) 812-815.

[14] H. Tochio, Q. Zhang, P. Mandal, M. Li, M. Zhang, Solution structure of the extended neuronal nitric oxide synthase PDZ domain complexed with an associated peptide [see comments]. Nat Struct Biol 6 (1999) 417-421.

[15] M.E. Kimple, J. Sondek, Affinity tag for protein purification and detection based on the disulfide-linked complex of InaD and NorpA. Biotechniques 33 (2002) 578, 580, 584-578 passim.

[16] H.E. Klock, E.J. Koesema, M.W. Knuth, S.A. Lesley, Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins 71 (2008) 982-994.

[17] H.E. Klock, S.A. Lesley, The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. Methods Mol Biol 498 (2009) 91-103.

[18] K. Wallenfels, Beta-galactosidase. Meth Enzymol 5 (1962) 212-218.

[19] K. Wallenfels, J. Lehmann, O.P. Malhotra, [Research on lactose-splitting enzymes. VII. The specificity of beta-galactosidase from E. coli ML 309]. Biochem Z 333 (1960) 209-225.

[20] G.R. Craven, E. Steers, Jr., C.B. Anfinsen, Purification, Composition, and Molecular Weight of the Beta-Galactosidase of Escherichia Coli K12. J Biol Chem 240 (1965) 2468-2477.

[21] W.V. Shaw, Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. Methods Enzymol 43 (1975) 737-755.

[22] E. Winshell, W.V. Shaw, Kinetics of induction and purification of chloramphenicol acetyltransferase from chloramphenicol-resistant Staphylococcus aureus. J Bacteriol 98 (1969) 1248-1257.

[23] G.V. Los, L.P. Encell, M.G. McDougall, D.D. Hartzell, N. Karassina, C. Zimprich, M.G. Wood, R. Learish, R.F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D.H. Klaubert, R.F. Bulleit, K.V. Wood, HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem Biol 3 (2008) 373-382.

[24] G.V. Los, K. Wood, The HaloTag: a novel technology for cell imaging and protein analysis. Methods Mol Biol 356 (2007) 195-208.

[25] T. Hata, M. Nakayama, Rapid single-tube method for small-scale affinity purification of polyclonal antibodies using HaloTag Technology. J Biochem Biophys Methods 70 (2007) 679-682.

[26] M.B. Kennedy, Signal-processing machines at the postsynaptic density. Science 290 (2000) 750-754.

[27] M.B. Kennedy, H.C. Beale, H.J. Carlisle, L.R. Washburn, Integration of biochemical signalling in spines. Nat Rev Neurosci 6 (2005) 423-434.

[28] M. Sheng, E. Kim, The postsynaptic organization of synapses. Cold Spring Harb Perspect Biol 3 (2011).

[29] C.A. Hunt, L.J. Schenker, M.B. Kennedy, PSD-95 is associated with the postsynaptic density and not with the presynaptic membrane at forebrain synapses. J Neurosci 16 (1996) 1380-1388.

[30] S. Nakanishi, Molecular diversity of glutamate receptors and implications for brain function. Science 258 (1992) 597-603.

[31] P.H. Seeburg, The molecular biology of mammalian glutamate receptor channels. Trends Neurosci 16 (1993) 359-365.

[32] M.L. Mayer, N. Armstrong, Structure and function of glutamate receptor ion channels. Annu Rev Physiol 66 (2004) 161-181.

[33] J.Z. Tsien, P.T. Huerta, S. Tonegawa, The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. Cell 87 (1996) 1327-1338.

[34] M.L. Mayer, Glutamate receptors at atomic resolution. Nature 440 (2006) 456-462.

[35] G. Kohr, V. Jensen, H.J. Koester, A.L. Mihaljevic, J.K. Utvik, A. Kvello, O.P. Ottersen, P.H. Seeburg, R. Sprengel, O. Hvalby, Intracellular domains of NMDA receptor subtypes are determinants for long-term potentiation induction. J Neurosci 23 (2003) 10791-10799.

[36] H.-J. Chen, M. Rojas-Soto, A. Oguni, M.B. Kennedy, A synaptic Ras-GTPase activating protein (p135 SynGAP) inhibited by CaM Kinase II. Neuron 20 (1998) 895-904.

[37] J.H. Kim, H.K. Lee, K. Takamiya, R.L. Huganir, The role of synaptic GTPase-activating protein in neuronal development and synaptic plasticity. J Neurosci 23 (2003) 1119-1124.

[38] L.E. Vazquez, H.J. Chen, I. Sokolova, I. Knuesel, M.B. Kennedy, SynGAP regulates spine formation. J Neurosci 24 (2004) 8862-8872.

[39] N.H. Komiyama, A.M. Watabe, H.J. Carlisle, K. Porter, P. Charlesworth, J. Monti, D.J. Strathdee, C.M. O'Carroll, S.J. Martin, R.G. Morris, T.J. O'Dell, S.G. Grant, SynGAP regulates ERK/MAPK signaling, synaptic plasticity, and

learning in the complex with postsynaptic density 95 and NMDA receptor. J Neurosci 22 (2002) 9721-9732.

[40] A.C. McMahon, M.W. Barnett, T.S. O'Leary, P.N. Stoney, M.O. Collins, S. Papadia, J.S. Choudhary, N.H. Komiyama, S.G. Grant, G.E. Hardingham, D.J. Wyllie, P.C. Kind, SynGAP isoforms exert opposing effects on synaptic strength. Nat Commun 3 (2012) 900.

[41] F.F. Hamdan, H. Daoud, A. Piton, J. Gauthier, S. Dobrzeniecka, M.O. Krebs, R. Joober, J.C. Lacaille, A. Nadeau, J.M. Milunsky, Z. Wang, L. Carmant, L. Mottron, M.H. Beauchamp, G.A. Rouleau, J.L. Michaud, De novo SYNGAP1 mutations in nonsyndromic intellectual disability and autism. Biol Psychiatry 69 (2011) 898-901.

[42] J.S. Oh, P. Manzerra, M.B. Kennedy, Regulation of the neuron-specific Ras GTPase activating protein, synGAP, by Ca2+/calmodulin-dependent protein kinase II. J Biol Chem 279 (2004) 17980-17988.

[43] V. Pena, M. Hothorn, A. Eberth, N. Kaschau, A. Parret, L. Gremer, F. Bonneau, M.R. Ahmadian, K. Scheffzek, The C2 domain of SynGAP is essential for stimulation of the Rap GTPase reaction. EMBO Rep 9 (2008) 350-355.

[44] M. Passafaro, C. Sala, M. Niethammer, M. Sheng, Microtubule binding by CRIPT and its potential role in the synaptic clustering of PSD-95. Nat Neurosci 2 (1999) 1063-1069.

[45] J.E. Brenman, J.R. Topinka, E.C. Cooper, A.W. McGee, J. Rosen, T. Milroy, H.J. Ralston, D.S. Bredt, Localization of postsynaptic density-93 to dendritic microtubules and interaction with microtubule-associated protein 1A. J Neurosci 18 (1998) 8805-8813.

[46] D.S. Bredt, S.H. Snyder, Isolation of nitric oxide synthetase, a calmodulin-requiring enzyme. Proc Natl Acad Sci USA 87 (1990) 682-685.

[47] J.E. Brenman, D.S. Bredt, Synaptic signaling by nitric oxide. Curr Opin Neurobiol 7 (1997) 374-378.

[48] D.S. Bredt, S.H. Snyder, Nitric oxide, a novel neuronal messenger. Neuron 8 (1992) 3-11.

[49] B.L. Firestein, J.E. Brenman, C. Aoki, A.M. Sanchez-Perez, A.E. El-Husseini, D.S. Bredt, Cypin: a cytosolic regulator of PSD-95 postsynaptic targeting. Neuron 24 (1999) 659-672.

[50] E.I. Charych, B.F. Akum, J.S. Goldberg, R.J. Jornsten, C. Rongo, J.Q. Zheng, B.L. Firestein, Activity-independent regulation of dendrite patterning by postsynaptic density protein PSD-95. J Neurosci 26 (2006) 10164-10176.

[51] B.F. Akum, M. Chen, S.I. Gunderson, G.M. Riefler, M.M. Scerri-Hansen, B.L. Firestein, Cypin regulates dendrite patterning in hippocampal neurons by promoting microtubule assembly. Nat Neurosci 7 (2004) 145-152.

[52] I.A. Lim, D.D. Hall, J.W. Hell, Selectivity and promiscuity of the first and second PDZ domains of PSD-95 and synapse-associated protein 102. J Biol Chem 277 (2002) 21697-21711.

[53] H. Morise, O. Shimomura, F.H. Johnson, J. Winant, Intermolecular energy transfer in the bioluminescent system of Aequorea. Biochemistry 13 (1974) 2656-2662.

[54] F.G. Prendergast, K.G. Mann, Chemical and physical properties of aequorin and the green fluorescent protein isolated from Aequorea forskalea. Biochemistry 17 (1978) 3448-3453.

[55] R.Y. Tsien, The green fluorescent protein. Annu Rev Biochem 67 (1998) 509-544.

[56] G. Miesenbock, D.A. De Angelis, J.E. Rothman, Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. Nature 394 (1998) 192-195.

[57] G.T. Hanson, T.B. McAnaney, E.S. Park, M.E. Rendell, D.K. Yarbrough, S. Chu, L. Xi, S.G. Boxer, M.H. Montrose, S.J. Remington, Green fluorescent protein variants as ratiometric dual emission pH sensors. 1. Structural

characterization and preliminary application. Biochemistry 41 (2002) 15477-15488.

[58] T.B. McAnaney, E.S. Park, G.T. Hanson, S.J. Remington, S.G. Boxer, Green fluorescent protein variants as ratiometric dual emission pH sensors. 2. Excited-state dynamics. Biochemistry 41 (2002) 15489-15494.

[59] M. Mank, A.F. Santos, S. Direnberger, T.D. Mrsic-Flogel, S.B. Hofer, V. Stein, T. Hendel, D.F. Reiff, C. Levelt, A. Borst, T. Bonhoeffer, M. Hubener, O. Griesbeck, A genetically encoded calcium indicator for chronic in vivo two-photon imaging. Nat Methods 5 (2008) 805-811.

[60] H. Ostergaard, A. Henriksen, F.G. Hansen, J.R. Winther, Shedding light on disulfide bond formation: engineering a redox switch in green fluorescent protein. EMBO J 20 (2001) 5853-5862.

[61] G.T. Hanson, R. Aggeler, D. Oglesbee, M. Cannon, R.A. Capaldi, R.Y. Tsien, S.J. Remington, Investigating mitochondrial redox potential with redox-sensitive green fluorescent protein indicators. J Biol Chem 279 (2004) 13044-13053.

[62] B.J. Baker, H. Mutoh, D. Dimitrov, W. Akemann, A. Perron, Y. Iwamoto, L. Jin, L.B. Cohen, E.Y. Isacoff, V.A. Pieribone, T. Hughes, T. Knopfel, Genetically encoded fluorescent sensors of membrane potential. Brain Cell Biol 36 (2008) 53-67.

[63] F.H. Wong, D.S. Banks, A. Abu-Arish, C. Fradin, A molecular thermometer based on fluorescent protein blinking. J Am Chem Soc 129 (2007) 10302-10303.

[64] A. Follenius-Wund, M. Bourotte, M. Schmitt, F. Iyice, H. Lami, J.J. Bourguignon, J. Haiech, C. Pigault, Fluorescent derivatives of the GFP chromophore give a new insight into the GFP fluorescence process. Biophys J 85 (2003) 1839-1850.

[65] G.S. Waldo, B.M. Standish, J. Berendzen, T.C. Terwilliger, Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol 17 (1999) 691-695.

[66] A. Kalnins, K. Otto, U. Ruther, B. Muller-Hill, Sequence of the lacZ gene of Escherichia coli. EMBO J 2 (1983) 593-597.

[67] A.V. Fowler, I. Zabin, Amino acid sequence of beta-galactosidase. XI. Peptide ordering procedures and the complete sequence. J Biol Chem 253 (1978) 5521-5525.

[68] R.E. Huber, M.N. Gupta, S.K. Khare, The active site and mechanism of the beta-galactosidase from Escherichia coli. Int J Biochem 26 (1994) 309-318.

[69] R.E. Huber, G. Kurz, K. Wallenfels, A quantitation of the factors which affect the hydrolase and transgalactosylase activities of beta-galactosidase (E. coli) on lactose. Biochemistry 15 (1976) 1994-2001.

[70] L. Guarente, M. Ptashne, Fusion of Escherichia coli lacZ to the cytochrome c gene of Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 78 (1981) 2199-2203.

[71] L. Guarente, Yeast promoters and lacZ fusions designed to study expression of cloned genes in yeast. Methods Enzymol 101 (1983) 181-191.

[72] A. Fire, S.W. Harrison, D. Dixon, A modular set of lacZ fusion vectors for studying gene expression in Caenorhabditis elegans. Gene 93 (1990) 189-198.

[73] J.T. Lis, J.A. Simon, C.A. Sutton, New heat shock puffs and beta-galactosidase activity resulting from transformation of Drosophila with an hsp70-lacZ hybrid gene. Cell 35 (1983) 403-410.

[74] C.V. Hall, P.E. Jacob, G.M. Ringold, F. Lee, Expression and regulation of Escherichia coli lacZ gene fusions in mammalian cells. J Mol Appl Genet 2 (1983) 101-109.

[75] J.H. Miller, A Short Course in Bacterial Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1992.

[76] J. Davies, F. Jacob, Genetic mapping of the regulator and operator genes of the lac operon. J Mol Biol 36 (1968) 413-417.

[77] A. Ullmann, F. Jacob, J. Monod, Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli. J Mol Biol 24 (1967) 339-343.

[78] M.E. Goldberg, S.J. Edelstein, Sedimentation equilibrium of paucidisperse systems. Subunit structure of complemented beta-galactosidase. J Mol Biol 46 (1969) 431-440.

[79] B. Muller-Hill, J. Kania, Lac repressor can be fused to beta-galactosidase. Nature 249 (1974) 561-563.

[80] A.V. Fowler, High-level production of beta-galactosidase by Escherichia coli merodiploids. J Bacteriol 112 (1972) 856-860.

[81] W.V. Shaw, The enzymatic acetylation of chloramphenicol by extracts of R factor-resistant Escherichia coli. J Biol Chem 242 (1967) 687-693.

[82] Y. Suzuki, S. Okamoto, The enzymatic acetylation of chloramphenicol by the multiple drug-resistant Escherichia coli carrying R factor. J Biol Chem 242 (1967) 4722-4730.

[83] C.M. Gorman, L.F. Moffat, B.H. Howard, Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol Cell Biol 2 (1982) 1044-1051.

[84] S.R. Kain, S. Ganguly, Overview of genetic reporter systems. Curr Protoc Mol Biol Chapter 9 (2001) Unit9 6.

[85] J.F. Thompson, L.S. Hayes, D.B. Lloyd, Modulation of firefly luciferase stability and impact on studies of gene regulation. Gene 103 (1991) 171-177.

FIGURES

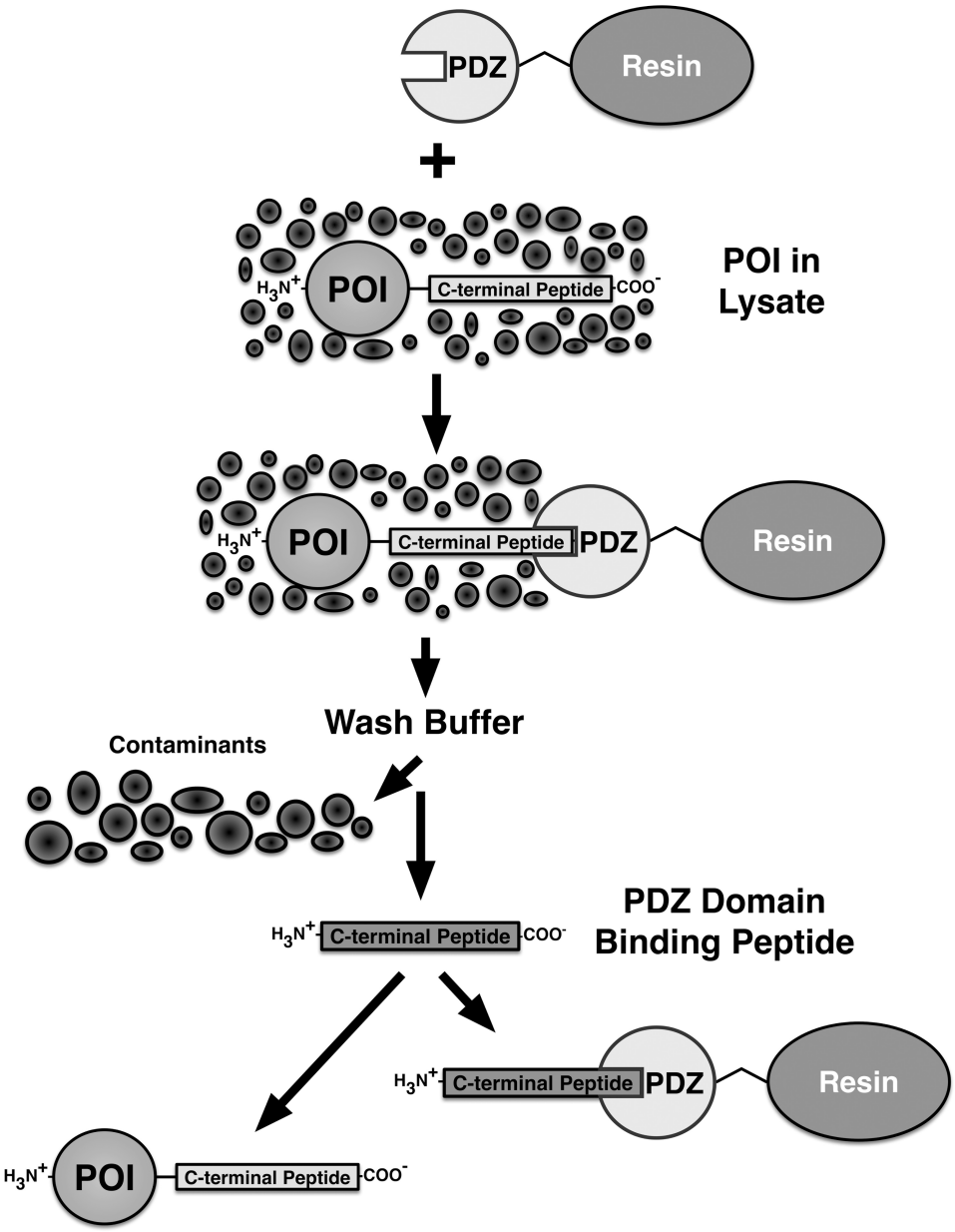


Figure 2.1. Generalized PDZ Domain Affinity Chromatography Purification Scheme I: PDZ Domain Peptide Ligand Tag. An expressed Protein Of Interest (POI), containing a naturally occurring or recombinantly appended PDZ Domain Peptide Ligand, can be captured from a cellular lysate by binding to a PDZ Domain coupled to a solid support resin. Extensive washing of the captured POI removes cellular contaminants. Free PDZ domain peptide ligand is then added to competitively elute the captured POI.

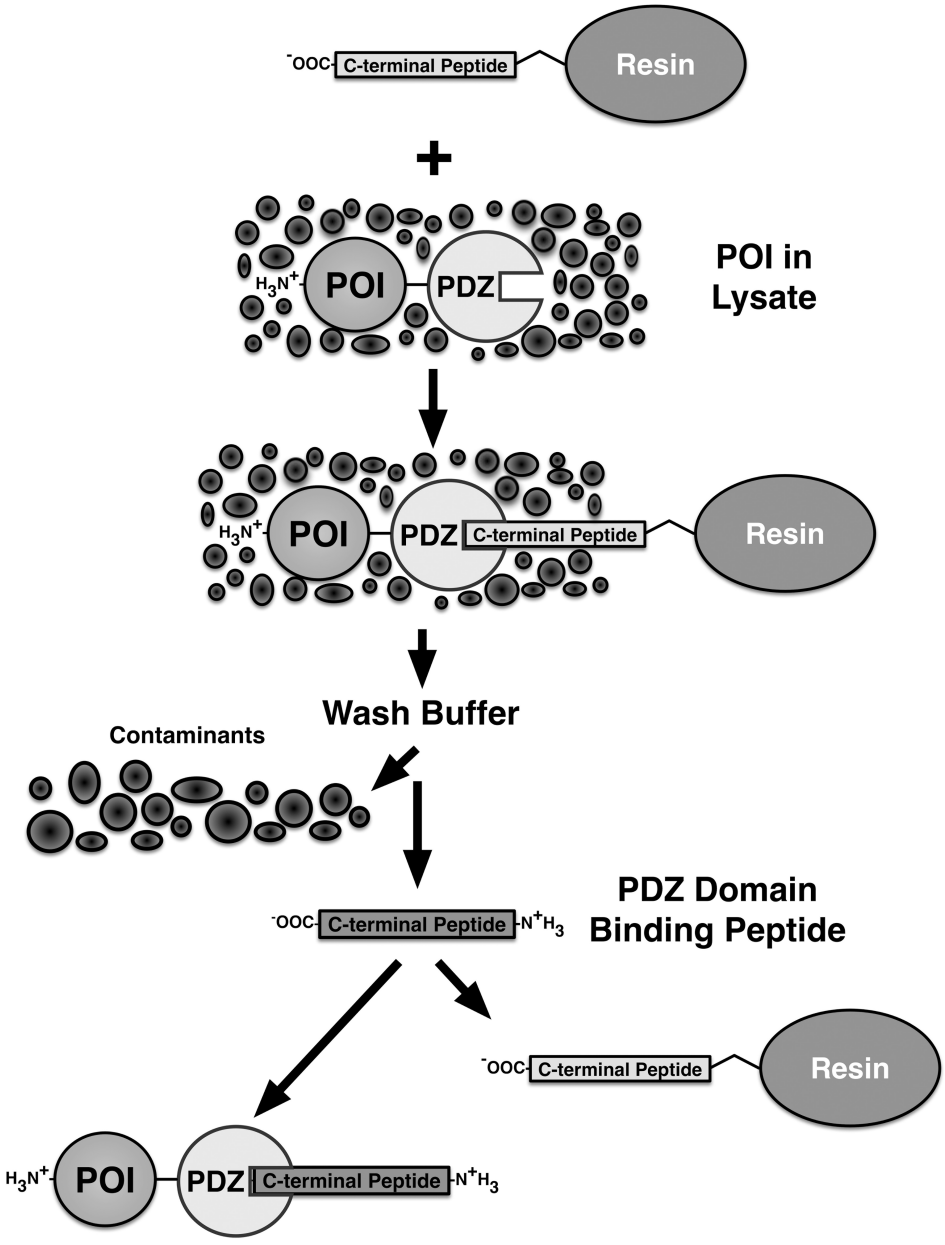


Figure 2.2. Generalized PDZ Domain Affinity Chromatography Purification Scheme II: PDZ Domain Tag. An expressed Protein Of Interest (POI) containing a naturally occurring or recombinantly appended PDZ domain can be captured from a cellular lysate by binding to a PDZ Domain-Peptide Ligand coupled to a solid support resin. Extensive washing of the captured POI removes cellular contaminants. As in Figure 1, free PDZ domain peptide ligand is then added to competitively elute the captured POI.

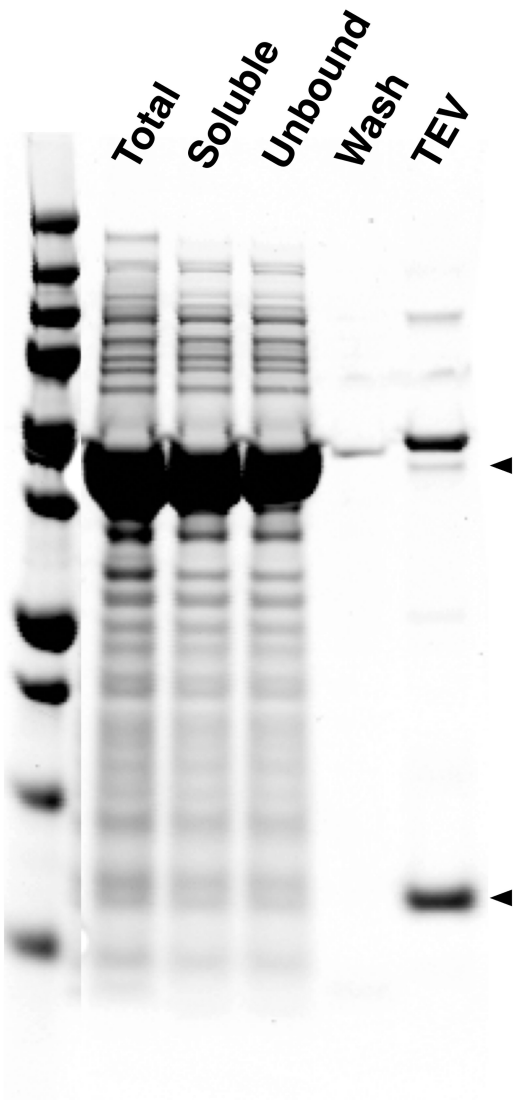


Figure 2.3. Preparation of PDZ Domain-HaloTag-HaloLink Affinity Resin. In the experiment shown here, total protein, soluble protein, unbound protein, wash and TEV protease-treated fractions from the preparation of PDZ3-HaloTag-HaloLink Resin were run on a 4-12% gradient SDS-PAGE gel with Precision Plus Protein All Blue Standards and stained with Gel Code Blue. A fusion protein in which the PDZ domain was fused to the HaloTag was prepared as described in Methods. The PDZ Domain Affinity Resin was synthesized by coupling of the fusion protein to HaloLink Resin. The derivatized resin was washed extensively to remove contaminants. To measure the binding capacity of the resin, bound PDZ domains were released from an aliquot of the resin by digestion with TEV protease, and quantified as described in Methods. In this experiment, resin capacity (pmol PDZ domain /µl resin) was 350 µM. Arrows indicate the molecular weight of PDZ3 domain-HaloTag (upper band) and free PDZ3 domain proteins (lower band).

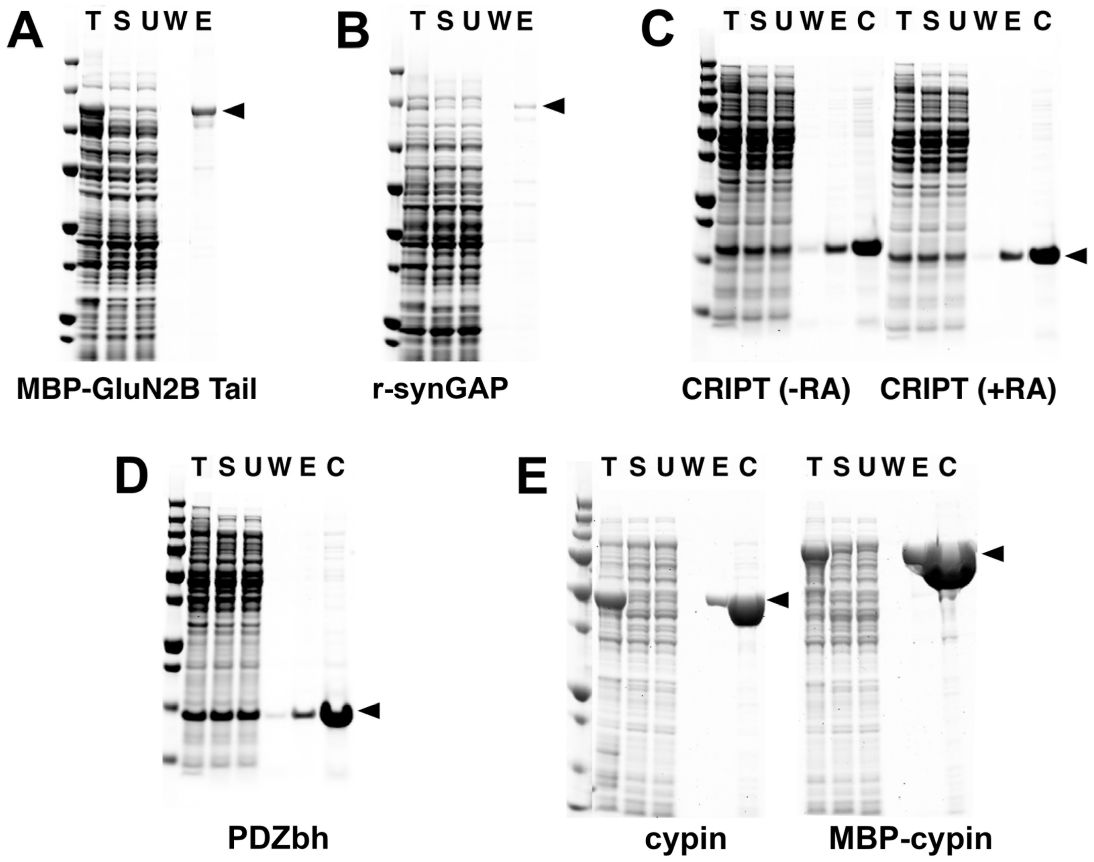


Figure 2.4. Purification of Heterologously Expressed Neuronal Proteins with Endogenous PDZ Domain Ligands. Six heterologously expressed neuronal proteins with naturally occurring PDZ domain ligands were purified on PDZ Domain Affinity Resins as described in Methods: (**A**) Purification of MBP-GluN2B Tail; (**B**) Purification of His-tagged r-synGAP; (**C**) Purification of CRIPT in the presence or absence of reducing agents (RA); (**D**) Purification of PDZbh; (**E**) Purification of cypin and MBP-cypin. MBP-GluN2B Tail and PDZbh were bound to PDZ2 Domain-HaloTag-HaloLink Resin and eluted with 400 μg/ml SIESDV peptide. R-synGAP and CRIPT were bound to PDZ3 Domain-HaloTag-HaloLink Resin and eluted with 400 μg/ml YKQTSV peptide. Cypin and MBP-cypin were bound to tandem PDZ1-PDZ2 Domain-HaloTag-HaloLink Resin and eluted with 400 μg/ml SIESDV peptide. Fractions from each purification and Precision Plus Protein All Blue Standards were separated on a 4-12% gradient SDS-PAGE gel and stained with Gel Code Blue. T, Total protein; S, soluble protein; U, unbound protein; W, wash; E, peptide eluate; C, concentrated peptide eluate.

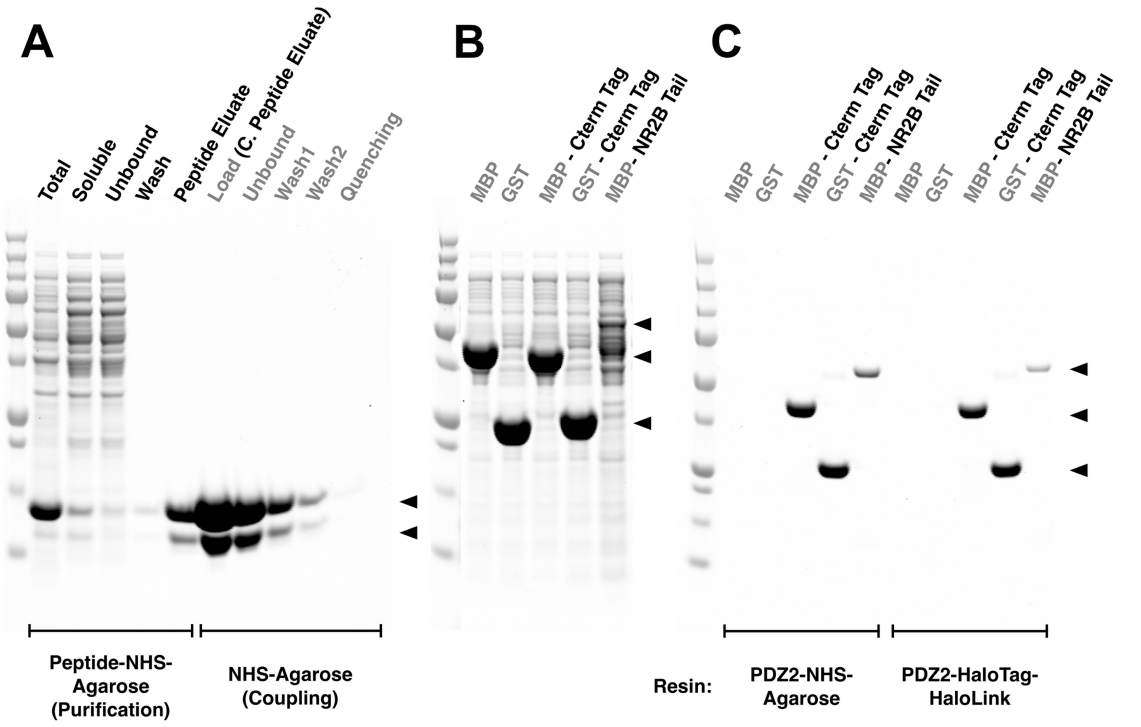


Figure 2.5. Comparison of HaloTag-HaloLink and NHS linkage for Synthesis of PDZ Domain-Agarose Affinity Resin. All samples and Precision Plus Protein All Blue Standards were fractionated by SDS-PAGE and protein was stained with Gel Code Blue. (**A**) Purification of PDZ2 and coupling to NHS Agarose resin. The first 5 lanes show fractions from the purification of the PDZ2 domain on GluN2B ligand-NHS-Agarose as described in Methods: Total, bacterial extract; Soluble, extract after centrifugation; Unbound, column flow-through; Wash; Peptide Eluate, protein eluted with 400 μg/ml SIESDV peptide. The peptide was removed from the PDZ2 domain, as described in Methods, before linkage to NHS-Agarose. The next 5 lanes show fractions from coupling of the purified PDZ2 Domain to NHS-Agarose beads (as described in Methods): Load, concentrated peptide eluate that was incubated with NHS-Agarose; Unbound, supernatant after linkage; Wash1 & 2, two washes of the resin before quenching; Quenching, supernatant after quenching of unreacted NHS groups with ethanolamine. Upper and lower arrows indicate full length and truncated PDZ2 domain, respectively. (**B**) Soluble lysate after heterologous expression of MBP, GST, their respective fusion proteins containing the 7-residue C-terminal PDZ2 Domain ligand Affinity Tag from GluN2B, and the entire GluN2B Tail fused to the C-terminus of MBP (see Methods). **C**) Purification of the fusion proteins shown in (B) on PDZ2-NHS-Agarose or on PDZ2 Domain-HaloTag-HaloLink resin. Soluble lysates were purified on the indicated Affinity Resins and eluted with 400 μg/ml SIESDV peptide, as described in Methods.

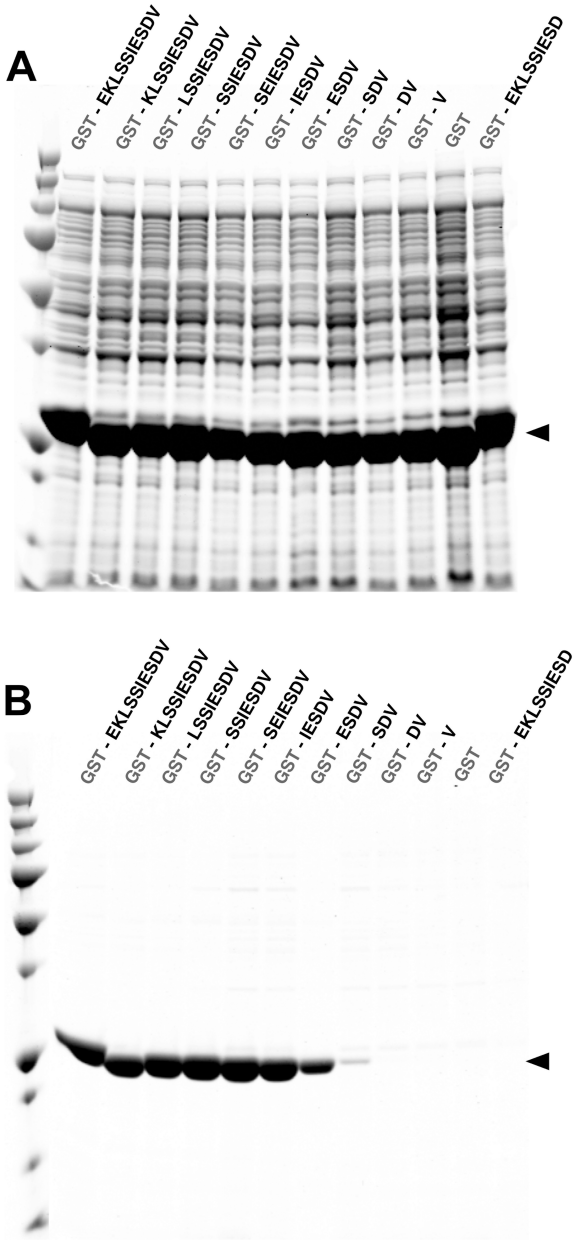


Figure 2.6. PDZ Domain Peptide Ligand Affinity Tags: Fusion of GluN2B Peptide Ligand Truncation Library to GST. All samples and Precision Plus Protein All Blue Standards were fractionated on a 4-12% gradient SDS-PAGE gel and stained with Gel Code Blue. (**A**) Soluble lysate after heterologous expression of Glutathione S-Transferase (GST) fused to a truncation library of the C-terminal ten residues of the GluN2B Tail. (**B**) Glutathione S-Transferase (GST) fusion proteins shown in (A), after purification on a PDZ2 Domain-HaloTag-HaloLink resin and elution with 400 μg/ml SIESDV peptide.

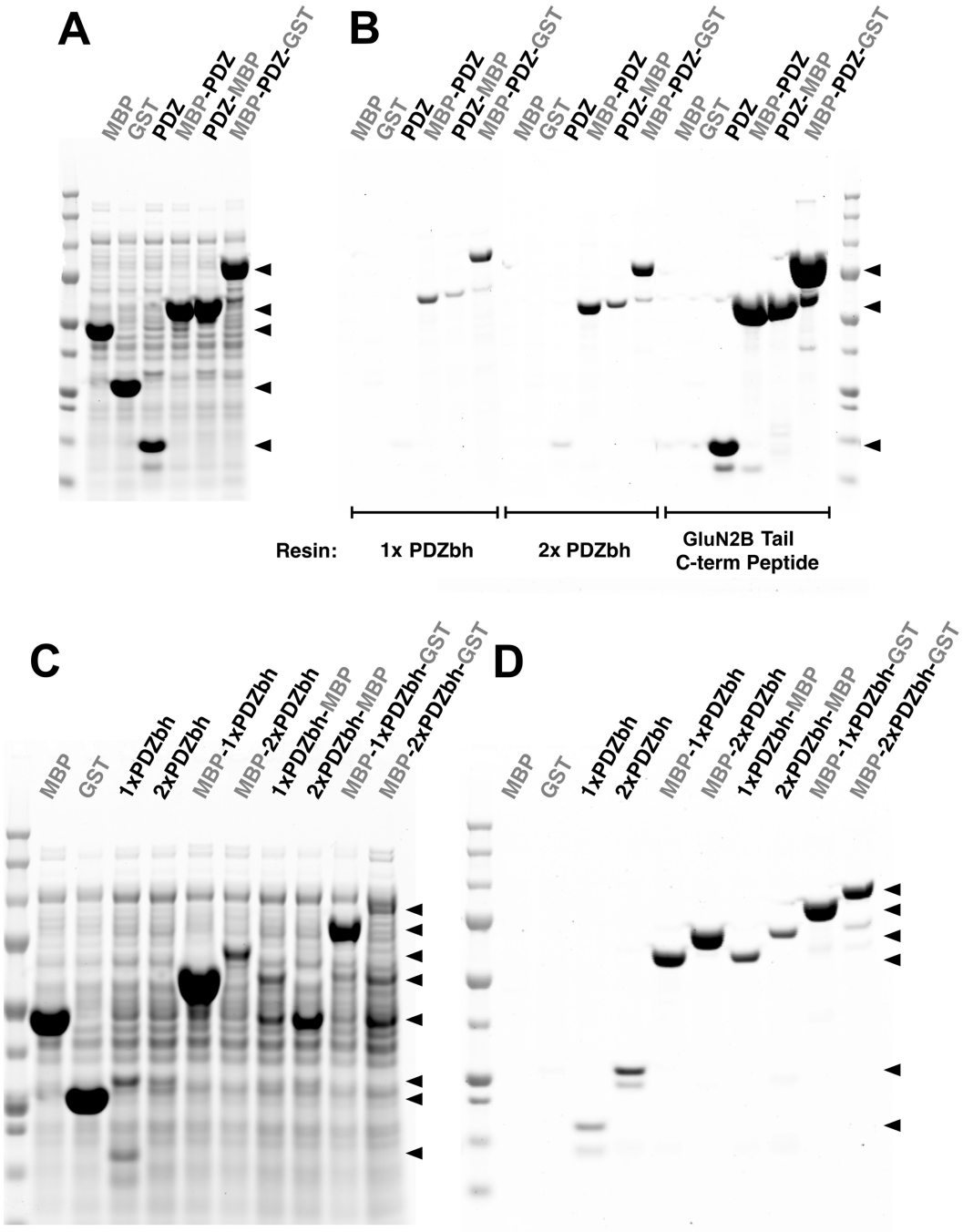


Figure 2.7. PDZ Domain and PDZ Domain-binding Beta Hairpin (PDZbh) Affinity Tags. All samples and Precision Plus Protein All Blue Standards were run on a 4-12% SDS-PAGE gradient gel and stained with Gel Code Blue. (**A**) Soluble lysates after heterologous expression of MBP, GST, a PDZ2 domain, MBP fusion proteins containing an N-terminal or C-terminal PDZ2 Tag, and an MBP-GST fusion protein containing an

internal PDZ2 Domain Tag. (**B)** Peptide eluates from purifications of the proteins shown in (A) on HaloTag-HaloLink Affinity Resins containing one (1x PDZbh) or two (2x PDZbh) PDZbh Domains, and on PDZ2 Domain Peptide Ligand-NHS-Agarose resin, prepared as described in Methods. Proteins were eluted with 400 μg/ml SIESDV peptide as described in Methods. (**C**) Soluble lysates after heterologous expression of MBP; GST; single and tandem PDZbh domains; and MBP, GST, and MBP-GST fusion proteins containing zero, one, or two N-terminal, C-terminal, or internal PDZbh domains. (**D**) Peptide eluates from purification of the proteins shown in (C) on PDZ2-HaloTag-HaloLink domain resin eluted with 400 μg/ml SIESDV peptide, as described in Methods.

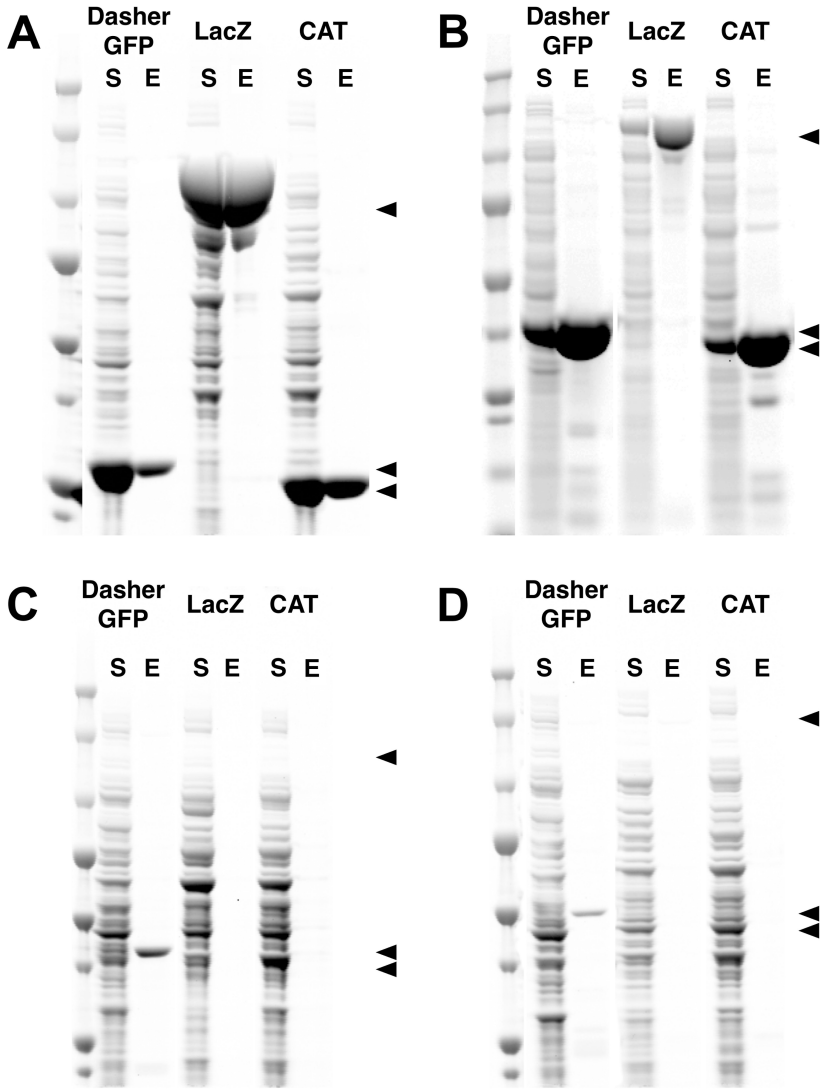


Figure 2.8. Purification of Proteins Using PDZ Domain and PDZ Peptide Ligand Affinity Tags. DasherGFP, LacZ and CAT were heterologously expressed with PDZ2 Domain or PDZ2 Peptide Ligand Tags and purified with the cognate Affinity Resin. For each protein, the soluble fraction (S) from the heterologous expression and the peptide eluate (E) after the purification were run on a 4-12% SDS-PAGE gradient gel with Precision Plus Protein All Blue Standards and stained with Gel Code Blue. (**A**) DasherGFP, LacZ and CAT were tagged at the C-terminus with the PDZ2 Domain Peptide Ligand, purified on PDZ2 Domain-NHS-Agarose Affinity Resin, and eluted with 400 µg/ml SIESDV peptide. (**B**) The three proteins were tagged at the N-terminus with a PDZ2 Domain, purified on PDZ2 Domain Peptide Ligand-NHS-Agarose resin, and eluted with 400 µg/ml SIESDV peptide. (**C**) The three proteins were tagged at the N-terminus with a PDZbh Domain or (**D**) two tandem PDZbh Domains, purified on PDZ2 Domain-NHS-Agarose Affinity Resin, and eluted with 400 µg/ml SIESDV peptide.

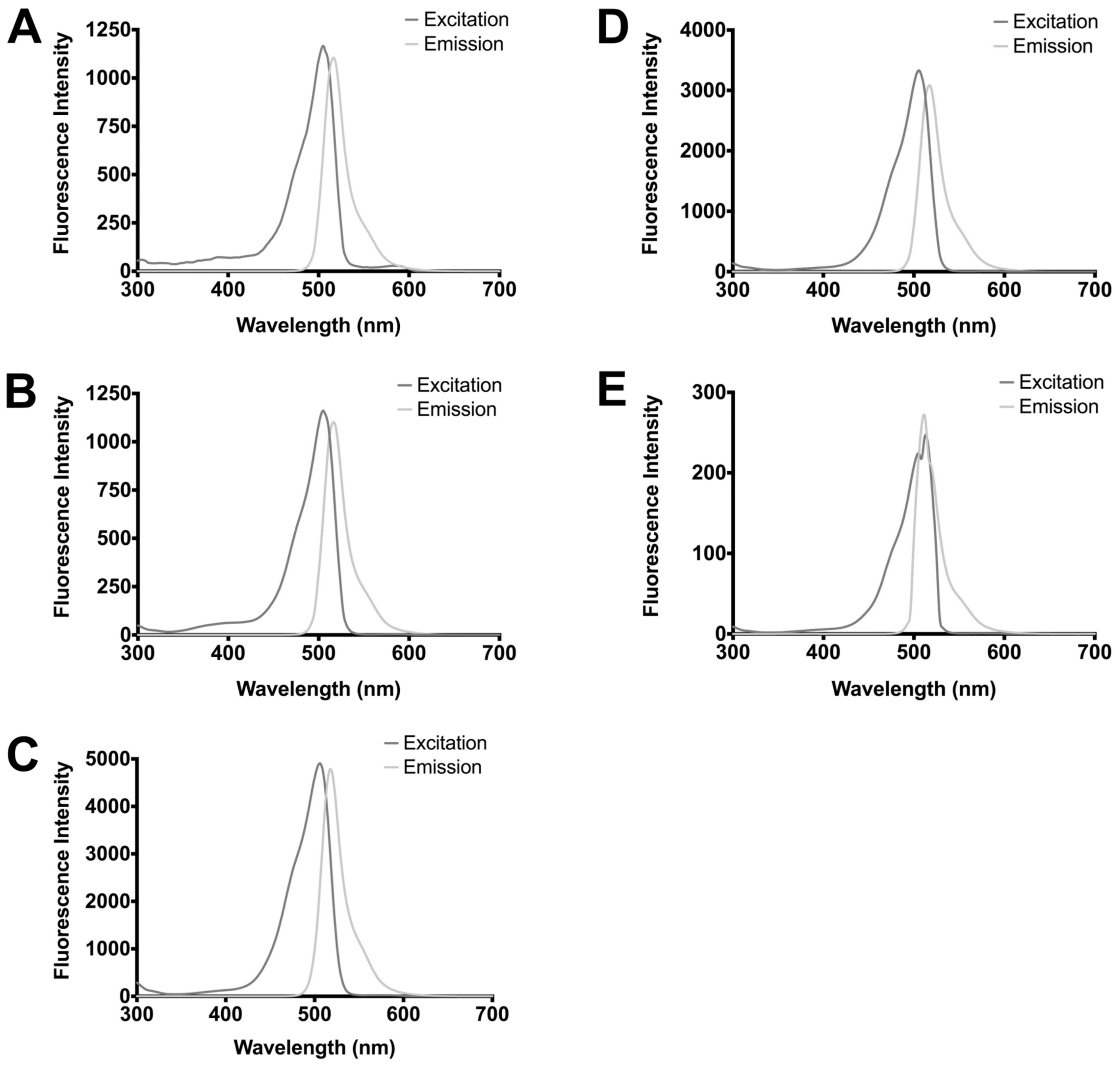


Figure 2.9. Functional Analysis of Purified DasherGFP Fusion Proteins. Fluorescence excitation and emission spectra of DasherGFP and DasherGFP fusion proteins were measured as described in Methods. (**A**) untagged; (**B**) fused to C-terminal PDZ Domain Peptide Ligand; (**C**) fused to N-terminal PDZ2 Domain; (**D**) fused to single PDZbh Domain; (**E**) fused to two tandem PDZbh Domains.

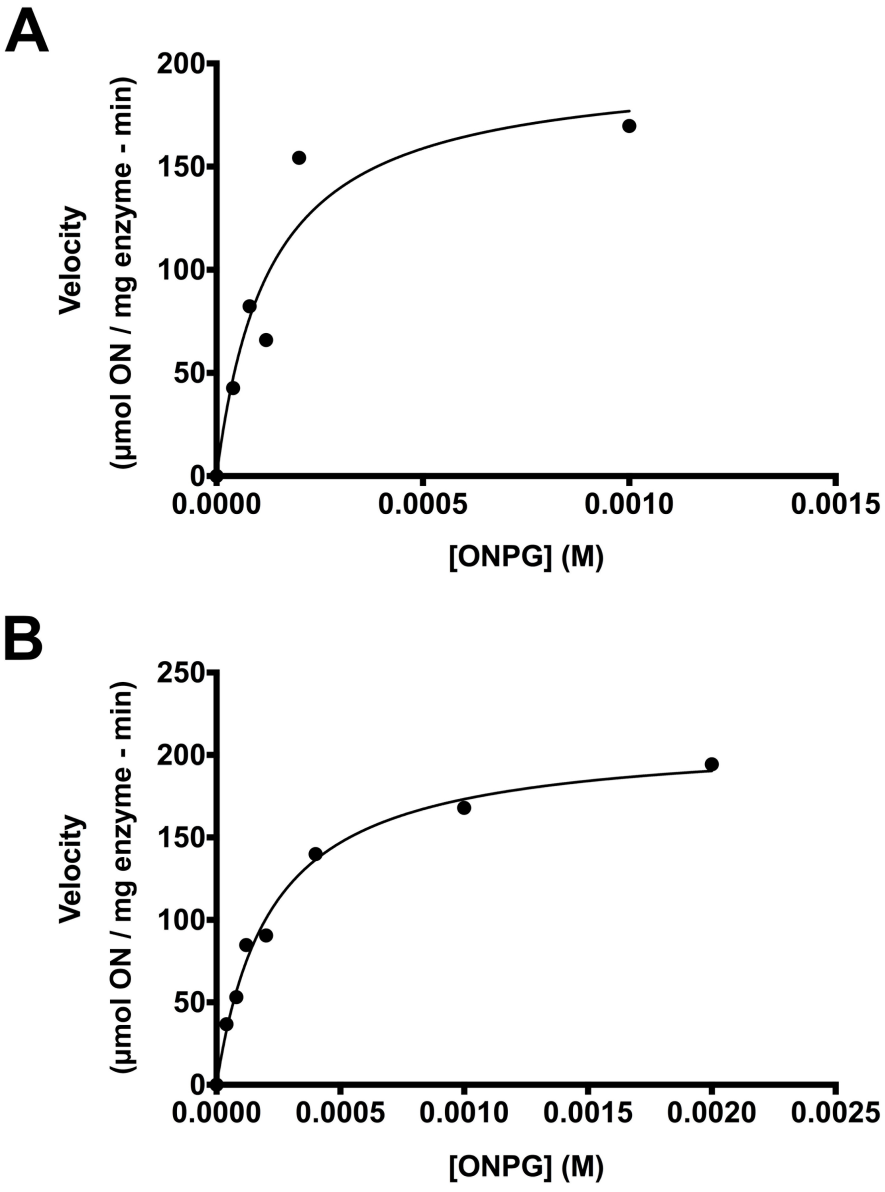


Figure 2.10. Functional Analysis of Purified LacZ Fusion Proteins. Michaelis-Menten plots for purified LacZ fused to (**A**) C-terminal PDZ2 domain peptide ligand; and (**B**) N-terminal PDZ2 domain. Galactosidase activity of LacZ was assayed by measuring hydrolysis of 0-2 mM ONPG for 10 minutes at 20 °C. Michaelis-Menten constants (Table 2.4) were calculated as described in Methods.

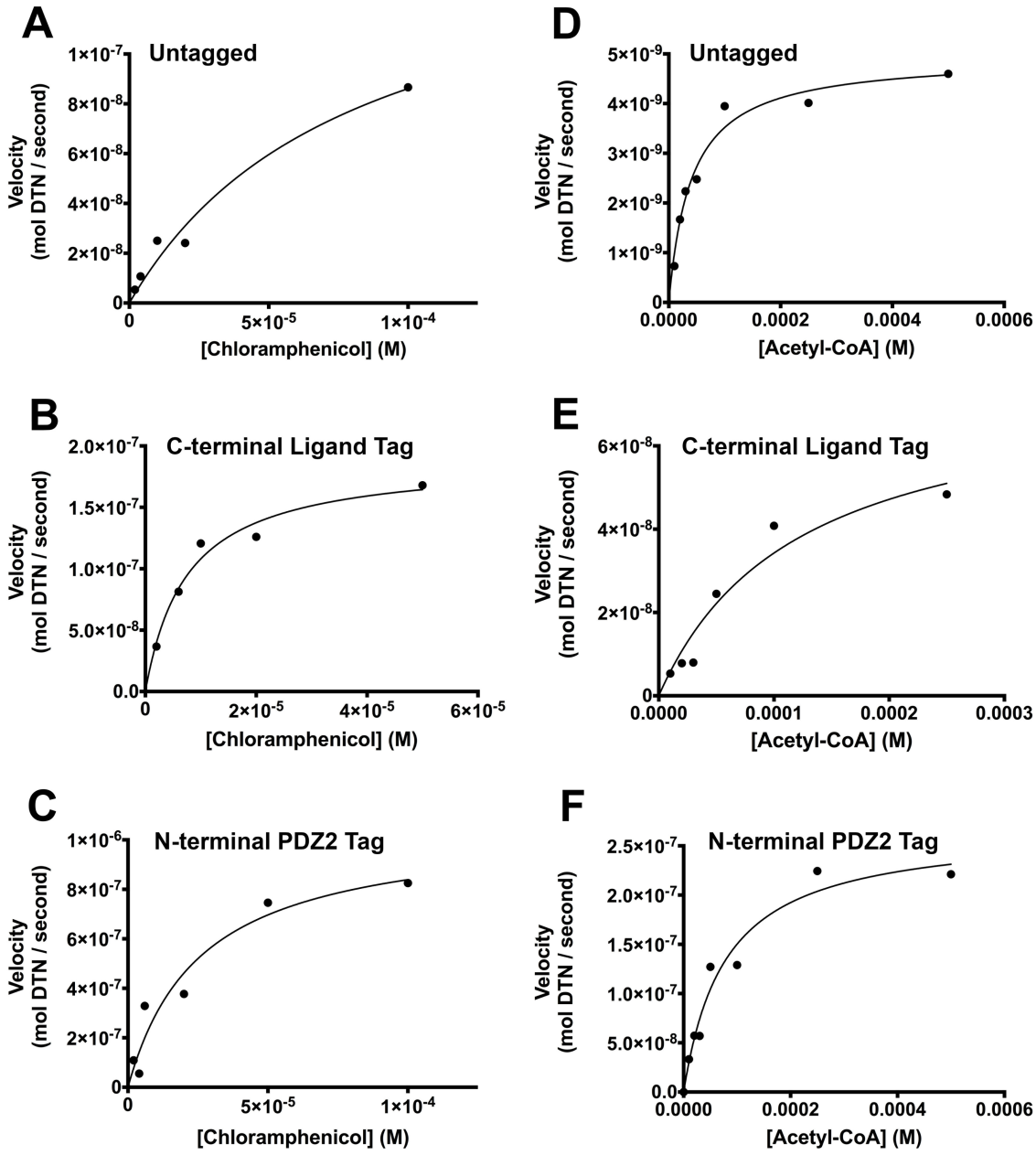


Figure 2.11. Functional Analysis of Purified Chloramphenicol Acetyl Transferase (CAT) Fusion Proteins. Michaelis-Menten plots for purified CAT: (**A** and **D**) untagged; (**B** and **E**) fused to a C-terminal PDZ2 domain peptide ligand; and (**C** and **F**) fused to an N-terminal PDZ2 domain. Transferase activity of purified CAT was assayed with (**A-C**) 500 μM Acetyl-CoA and 0-100 μM Chloramphenicol, or with (**D-F**) 100 μM Chloramphenicol and 0-500 μM Acetyl-CoA for 10 minutes at 37 °C. Michaelis-Menten constants (Table 2.5) were calculated as described in Methods.

TABLES

Table 2.1. Heterologously Expressed Neuronal Proteins Containing Endogenous PDZ Domain Ligands

Name	Organism	Residues	N-terminal Tag	PDZ Domain Resin	Elution Peptide
GluN2B Tail	*Mus musculus*	842-1482	MBP	PDZ2	SIESDV
r-synGAP	*Rattus norvegicus*	103-1293	His Tag	PDZ3	YKQTSV
CRIPT	*Rattus norvegicus*	Full Length	-	PDZ3	YKQTSV
nNOS PDZbh	*Mus musculus*	11-129	-	PDZ2	SIESDV
cypin	*Homo sapiens*	Full Length	MBP or His Tag	PDZ1+PDZ2	SIESDV

Table 2.2. Ligand Densities of PDZ Domain-HaloTag-HaloLink Affinity Resins

Ligand densities on the resin were measured as described in Methods.

PDZ Domain Resin	Ligand Density (pmol PDZ Domain / µl Resin)
PDZ1+PDZ2	*78*
PDZ1	*100*
PDZ2	*266*
PDZ3	*350*
Single PDZbh	*89.4*
Tandem PDZbh	*45*

Table 2.3. Functional Analysis of Dasher GFP and Purified DasherGFP Fusion Proteins

Fluorescence emission at 520 nm, relative to untagged DasherGFP, of equimolar amounts of purified DasherGFP fused to a C-terminal PDZ2 Domain Peptide Ligand, an N-terminal PDZ2 Domain, and single or tandem PDZbh Domains.

Affinity Tag	Relative Fluorescence at 520 nm (%)
Untagged DasherGFP	100
C-terminal PDZ Domain Peptide Ligand	99.2
N-terminal PDZ Domain	97.5
N-terminal PDZ Domain+Beta Hairpin	90.1
N-terminal 2X PDZ Domain+Beta hH	0.3

Table 2.4. Michaelis-Menten Constants of LacZ and Purified LacZ Fusion Proteins

Comparison of k_{cat} and K_M values for LacZ and purified LacZ fused to a C-terminal PDZ2 Domain Peptide Ligand or an N-terminal PDZ2 Domain.

Affinity Tag	k_{cat} (sec^{-1})	K_M (μM)
None	345*	161*
C-terminal PDZ Domain Peptide Ligand	413±38	136±44
N-terminal PDZ Domain	390±10	227±21

*Wallenfels K. and Malhotra P. (1962) Advanced Carbohydrate Chemistry: 239-298

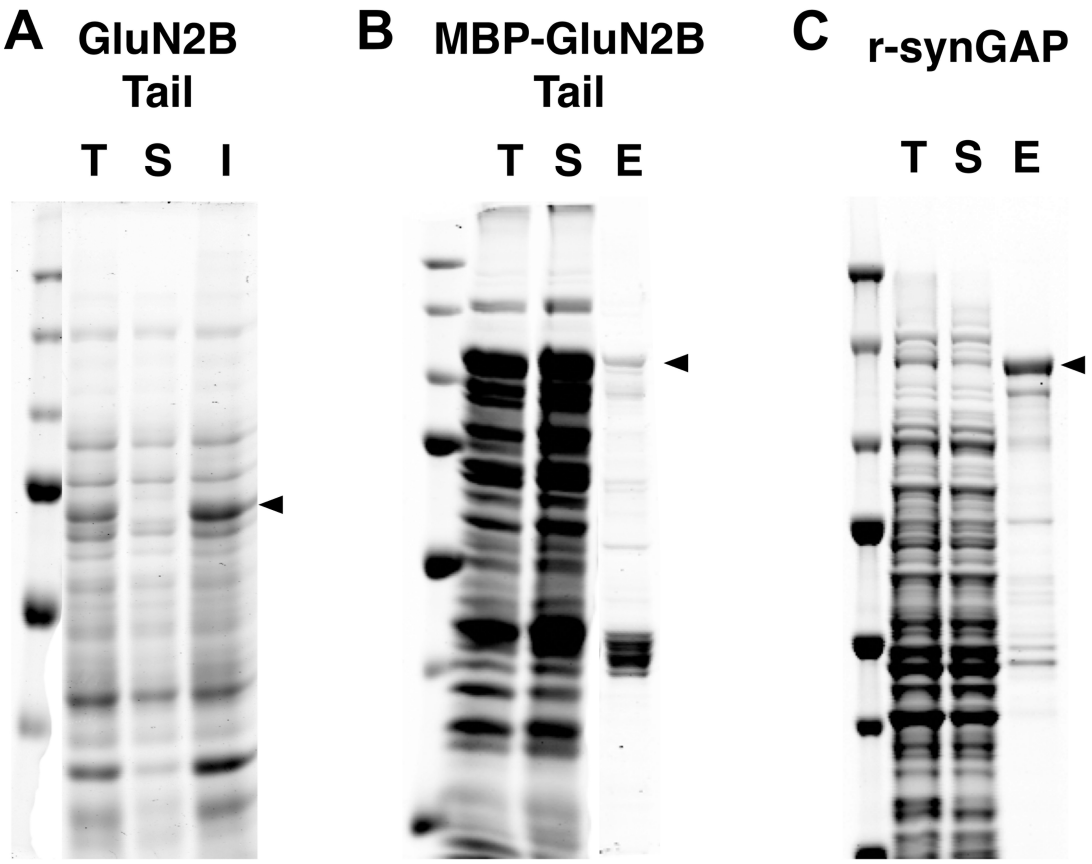
Table 2.5. Michaelis-Menten Constants of CAT and Purified CAT Fusion Proteins

Comparison of k_{cat} and K_M values for CAT and purified CAT fused to a C-terminal PDZ2 Domain Peptide Ligand or an N-terminal PDZ2 Domain.

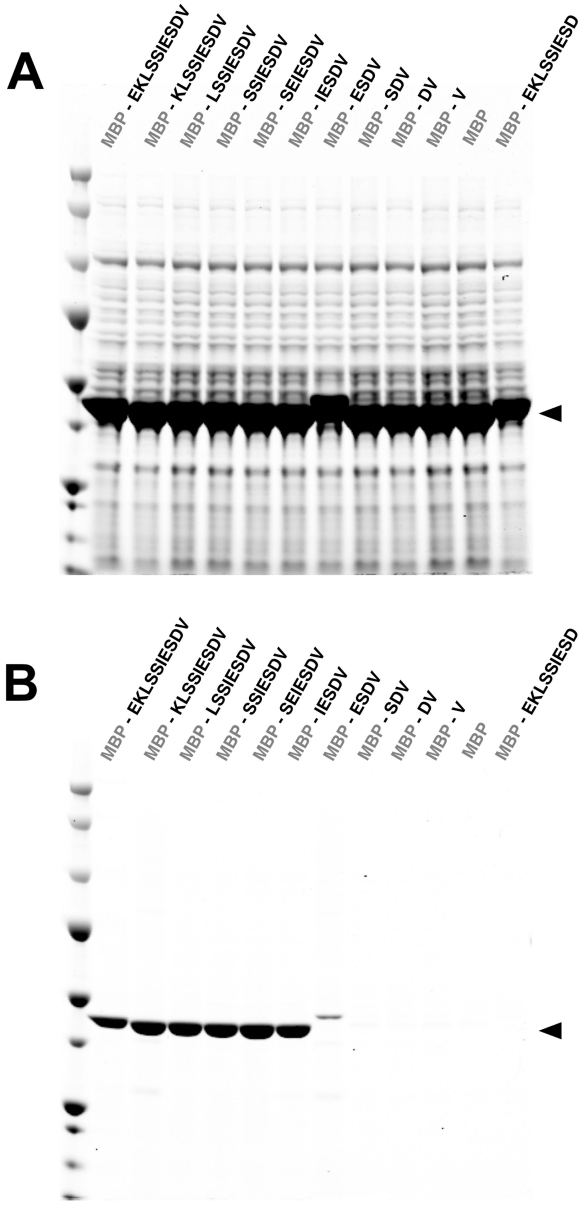
Affinity Tag	Chloramphenicol**		Acetyl-CoA***	
	k_{cat} (sec^{-1})	K_M (μM)	k_{cat} (sec^{-1})	K_M (μM)
None*	0.43 ± 0.11	80 ± 40	0.085 ± 0.004	41 ± 7
C-terminal PDZ Domain Peptide Ligand	0.37 ± 0.03	8 ± 2	0.074 ± 0.02	120 ± 50
N-terminal PDZ Domain	0.37 ± 0.06	26 ± 10	0.093 ± 0.008	78 ± 20

*CAT was purchased from Sigma Aldrich (Cat. No. C8413, St. Louis MO).
**Chloramphenicol concentration ranged from 0-100 μM. Acetyl-CoA concentration was fixed at 500 μM.
***Acetyl-CoA concentration ranged from 0-500 μM. Chloramphenicol concentration was fixed at 100 μM.

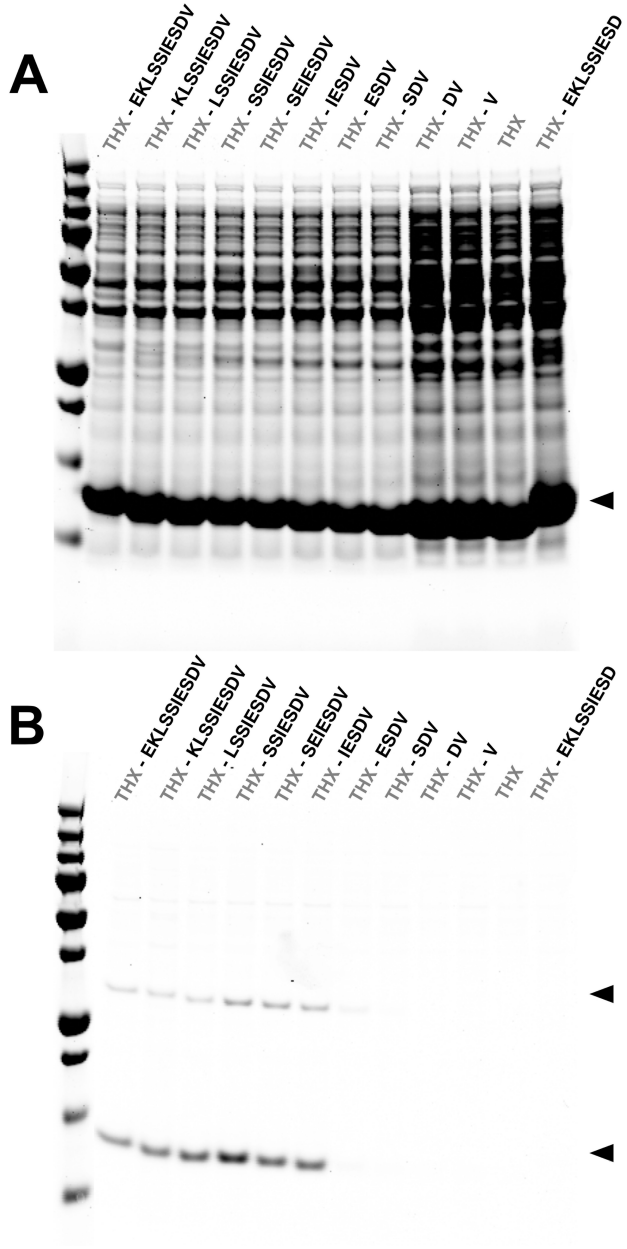
SUPPLEMENTAL FIGURES



Supplemental Figure 2.1: Expression of MBP-GluN2B Tail and Purification of MBP-GluN2B Tail and R-synGAP on Commercially Available Affinity Chromatography Systems. Fractions from attempted expression and purification of proteins on commercially available Affinity Resins and Precision Plus Protein All Blue Standards were separated on 4-12% gradient SDS-PAGE gels and stained with Gel Code Blue. (**A**) The GluN2B Tail expressed without a tag was insoluble after expression in *E. coli*. T, total bacterial lysate; S, soluble bacterial protein; and I, insoluble bacterial protein pelleted by centrifugation. (**B**) The GluN2B Tail expressed as an N-terminal MBP fusion protein (MBP-GluN2B Tail) was soluble and could be purified with amylose affinity resin (E8021; NEB, Ipswitch MA) and eluted with 10 mM maltose. The product was approximately 8% pure and was recovered at a concentration of 0.1 μM with a yield of 6%. T, total protein; S, soluble protein; and E, eluate. (**C**) R-synGAP was purified with Talon Cobalt-CMA affinity resin (Cat. No. 635503; Clontech Laboratories, Inc., Mountain View CA) and eluted with 10 mM EDTA. The product was approximately 25% pure and was recovered at a concentration of 2 μM with a yield of 75%. T, total protein; S, soluble protein; and E, eluate.



Supplemental Figure 2.2. Development of PDZ Domain Ligand Affinity Tags: Fusion of GluN2B C-terminal Peptide Truncation Library to MBP. Heterologous expression and purification of MBP fused to a truncation library of the C-terminal ten residues of the GluN2B Tail. All samples and Precision Plus Protein All Blue Standards were fractionated on a 4-12% gradient SDS-PAGE gel and stained with Gel Code Blue. (**A**) Clarified lysate after bacterial expression. (**B**) Eluates from purification of MBP-GluN2B Tail C-terminal truncations on PDZ2-HaloTag-HaloLink Resin, eluted with 400 μg/ml SIESDV peptide.



Supplemental Figure 2.3. Development of PDZ Domain Ligand Affinity Tags: Fusion of GluN2B C-terminal Peptide Truncation Library to THX. Heterologous expression and purification of THX fused to a truncation library of the C-terminal ten residues of the GluN2B Tail. All samples and Precision Plus Protein All Blue Standards were fractionated on a 4-12% gradient SDS-PAGE gel and stained with Gel Code Blue. (**A**) Clarified lysate after bacterial expression. (**B**) Eluate from purification of THX-GluN2B Tail C-terminal truncations on PDZ2-HaloTag-HaloLink Resin, eluted with 400 µg/ml SIESDV peptide.

Chapter III

DIFFERENTIAL REGULATION OF RAS AND RAP GAP ACTIVITIES OF SYNGAP BY THE PROTEIN KINASES CAMKII AND CDK5

AUTHORS

Ward G. Walkup IV, Robert L. Graham[#], Mike J. Sweredoski[#], Lorraine Washburn[#], Sonja Hess, Mary B. Kennedy*

Division of Biology and Biological Engineering, California Institute of Technology, 1200 East California Blvd, Mail Code 216-76, Pasadena CA 91125, USA

*To whom correspondence should be addressed. Tel: +1 6263953923; Fax: +1 6263958474; Email: kennedym@its.caltech.edu

Key Words:

synGAP, CDK5, CaMKII, GTPase, Phosphorylation, Synaptic Plasticity, Mass Spectrometry

Author Contributions:

WGWIV, LW, MJS, and RLG conceived, designed, and carried out experiments; analyzed and interpreted data; and wrote the paper. SH and MBK conceived experiments, interpreted data, and wrote the paper.

AUTHOR'S COMMENTS

This Chapter is a draft of a manuscript destined for submission to the Journal of Biological Chemistry in early 2014. While data collection is nearly complete, several small experiments required for completion of this manuscript are absent from this chapter (see Additional Experiments to Complete Before Manuscript Submission). Once this additional data has been collected, the Results and Discussion section will be significantly expanded in scope.

ABSTRACT

synGAP is a neuron-specific Ras and Rap GTPase-activating protein (GAP) found in high concentration in the postsynaptic density fraction from mammalian forebrain. Proteins in the postsynaptic density, including synGAP, are part of a signaling complex attached to the cytoplasmic tail of the N-methyl-D-aspartate-type glutamate receptor (NMDAR). synGAP bound to the postsynaptic density, or Sf9 cell membrane preparations, has previously been shown to be phosphorylated by a prominent component of the NMDAR complex, Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII). Here we show that phosphorylation of soluble, purified recombinant synGAP (r-synGAP) by the protein kinases CaMKII and cyclin-dependent kinase 5 (CDK5) increases its HRas and Rap1 GAP activity, but has no effect on its Rap2 GAP activity. Phosphorylation of r-synGAP by CaMKII increases its HRas and Rap1 GAP

activities by 25% and 76%, respectively, whereas phosphorylation by CDK5 increases its HRas and Rap1 GAP activities by 65% and 20%, respectively. Using mass spectrometry, we confirm previously identified CaMKII phosphorylation sites in synGAP (S751, S765, S1093, S1099, and S1123) and identify novel CaMKII (S737, S821, S825, S827, S882, T897, S1150, S1210, and S1283) and CDK5 phosphorylation sites (S728, S773, S802, S842, and S1093). CDK5 primarily phosphorylates synGAP at two main sites, S773/T775 and S802. Activity assays of phosphomimetic synGAP mutants suggest that phosphorylation at S773/T775 inhibits synGAP activity, whereas phosphorylation at S802 increases synGAP activity. These phosphomimetic mutants also suggest that the net effect of phosphorylation at both the S773/T775 and S802 sites is r-synGAP activation.

INTRODUCTION

SynGAP is a dual Ras and Rap GTPase activating protein (GAP) that is highly concentrated in the postsynaptic density (PSD) of excitatory synapses [1-3]. synGAP is positioned in close proximity to the cytoplasmic tail regions of the NMDAR through its association with the PDZ domains of PSD-95 [1, 2]. In spite of synGAP's homology to the RasGAPs, p120GAP, and neurofibromin, synGAP stimulates the GTPase activity of Rap much more potently than of Ras [3, 4]. Like other small GTPases, Ras and Rap adopt an active conformation when bound to GTP and are inactivated by GTP hydrolysis. GAPs

catalyze the inactivation of small GTPases by accelerating their rate of endogenous GTP hydrolysis. Influx of Ca^{2+} through the N-methyl-D-aspartate-type glutamate receptors (NMDARs) leads to the the activation of the small GTPases Ras and Rap. Differential activation of Ras and Rap acts to modulate synaptic strength by promoting the insertion (exocytosis) or removal (endocytosis), respectively, of 2-amino-3-(3-hydroxy-5-methyl-isoxazol-4-yl)propanoic acid receptors (AMPARs) from the synapse [5]. We investigated whether phosphorylation of individual residues on synGAP differentially regulates its Ras and Rap GAP activities, such that the influence of synGAP on inactivation of these two small GTPases varies under different physiological conditions.

We previously showed that activation of NMDARs leads to phosphorylation of synGAP by CaMKII [1, 6], and we identified several major sites on synGAP, that when phosphorylated by CaMKII, increase its Ras GAP activity [6]. Here we show that phosphorylation of purified, soluble recombinant synGAP (r-synGAP) by CaMKII increases its Ras and Rap1 GAP activity by 25% and 76%, respectively, but does not affect its Rap2 GAP activity. Thus, phosphorylation of synGAP by CaMKII accelerates the rate of inactivation of Rap1 more potently than of HRas, and may lead to alteration of the ratio of activated Ras and Rap in synapses. Using mass spectrometry, we confirmed previously identified CaMKII phosphorylation sites in r-synGAP (S751, S765, S1093, S1099, and S1123) and identified novel CaMKII phosphorylation sites (S737, S821, S825, S827, S882, T897, S1150, S1210, and S1283). We then used mass spectrometric analysis to identify a CaMKII-independent site of phosphorylation on PSD bound synGAP (S773), and determined for the first time that synGAP is a substrate for cyclin-dependent kinase 5

(CDK5), which phosphorylates it primarily at S773. CDK5 is a proline directed serine/threonine kinase that is important for many neuronal processes. CDK5 has previously been implicated in the regulation of synaptic plasticity [7, 8], localized to synapses, and shown to interact with or regulate many synaptic proteins, including NR2A [9], PSD-95 [10], and CaMKII [11].

Using mass spectrometry, we confirm that r-synGAP is phosphorylated by CDK5 at sites S728, S773, S802, S842, and S1093. Of these sites, CDK5 primarily phosphorylates r-synGAP at two main sites, S773/T775 and S802. We show that phosphorylation of r-synGAP by CDK5 increases its HRas and Rap1 GAP activities by 65% and 20%, respectively, and may lead to alteration of the ratio of activated Ras and Rap in synapses. Activity assays using phosphomimetic r-synGAP mutants suggest that phosphorylation at S773/T775 inhibits r-synGAP activity, whereas phosphorylation at S802 increases r-synGAP activity. r-synGAP phosphomimetic mutants also suggest that the net effect of phosphorylation at both S773/T775 and S802 sites is synGAP activation.

MATERIALS AND METHODS

Preparation of Postsynaptic Density from Rat Brain. The PSD fraction was prepared as described previously *[12]* by a modification of the method of Carlin *et al*. *[13]*. We were unable to extract synGAP from the PSD fraction without destroying its GAP activity. To ascertain the exact contribution of CDK5 and CaMKII phosphorylation

on synGAP activity, we purified soluble, recombinant synGAP (r-synGAP) from *E.coli* (as described below).

Mass Spectrometry and Identification of CDK5 Phosphorylation Site in synGAP. Liquid chromatography-tandem mass spectrometry (LC/MS/MS) was conducted by the California Institute of Technology Protein and Peptide Microanalytical Laboratory. Rat PSD fraction (65 µg) was phosphorylated for 2 minutes at 30°C in a 30 µl reaction mix (Ca/CaM) containing 50 mM Tris-HCl (pH 8.0), 10 mM MgCl2, 0.4 mM EGTA, and 10 mM DTT, prewarmed for 2 minutes at 30°C and initiated with the addition of 30 µM ATP. Phosphorylated PSD proteins were fractionated by SDS-PAGE, stained with Coomassie blue, and the 135 kDa band containing synGAP was excised and digested with trypsin in the gel as described in *[14]*. Phosphopeptides in the digests were identified by nanoflow capillary high performance liquid chromatography coupled with triple quadrupole/electrospray tandem mass spectrometry (Nano LC-2D, Eksigent; QSTARXL, Applied Biosystems). Data from the MS/MS spectra were analyzed with the Mascot MS/MS Ions search (Matrix Science), to identify phospho-peptides from synGAP. The search parameters were set as follows: data base=NCBInr; taxonomy=Rattus; variable modification of phosphorylation on serine, threonine, and tyrosine; peptide tolerance= ±0.5 Da; MS/MS tolerance= ±0.08 Da.

Cloning, Expression and Purification of r-synGAP. *G*ene fragments coding for residues 103-1293 and 103-725 of rat synGAP (AF048976) were PCR amplified with

PfuUltra II Fusion HS (#600670, Agilent) and cloned by PIPE-LIC *[15, 16]* into pET-47b(+) (#71461, EMD Millipore) in frame with an N-terminal 6x Histidine Tag and a PreScission Protease cleavage site. Sequence verified plasmids were transformed into the Rosetta2(DE3) *E. coli* strain (#71397, EMD Millipore) for protein expression. Single colonies were grown overnight at 37 °C in 5 ml of lysogeny broth (LB) (#L9110, Teknova) supplemented with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol. Overnight cultures were diluted 1:500 into LB and grown at 37 °C for ~4-6 hours to an $O.D._{600}$ of 1.0. Cultures were then chilled to 18°C, and protein expression was induced with 0.2 mM IPTG for 24 hours. Bacterial pellets were harvested by centrifugation, and resuspended in Lysis/Wash Buffer (20 mM Tris, 500 mM NaCl, 10 mM TCEP, 5 mM $MgCl_2$, 60 mM Imidazole, 2 mM ATP, 1 mM PMSF, 0.2% Tergitol Type NP-40, and Complete EDTA free protease inhibitor, pH 7.0) supplemented with 25 U/ml Benzonase (#71206, EMD Millipore) and 10 kU/ml ReadyLyse (#R1810M, Epicentre). Cells were lysed by microfluidization (#ML-110, Microfluidics), and insoluble material was removed by centrifugation at 30,000 x g. Supernatant was loaded onto Talon Cobalt-CMA resin (#635503, Clontech) via a hybrid batch gravity procedure, washed with Lysis/Wash Buffer, and eluted using Lysis/Wash Buffer (lacking 2 mM ATP) supplemented with 10 mM EDTA and 250 mM Imidazole. Fractions containing synGAP were pooled, and concentrated using ultrafiltration (#89923 for 103-1293, #89887A for 103-725; Pierce). Concentrated samples of 103-1293 (0.25-1 mg/ml) were then buffer exchanged by ultrafiltration into Storage Buffer (20 mM Tris, 500 mM NaCl, 10 mM TCEP, 5 mM $MgCl_2$, 1mM PMSF, 0.2% Tergitol Type NP-40, and Complete

EDTA free protease inhibitor, pH 7.0), flash frozen in liquid nitrogen, and stored at -80°C. Concentrated samples of 103-725 were further purified on a Hi Load 26/600 Superdex 200 prep grade column (#28-9893-36, GE Healthcare) equilibrated with Storage Buffer. Fractions containing synGAP were pooled, concentrated by ultrafiltration to 2 mg/ml, flash frozen in liquid nitrogen, and stored at -80°C. All 103-1293 synGAP mutants were purified with the same procedure used for wild type 103-1293 synGAP.

Cloning, Expression and Purification of HRas, Rap1 and Rap2. Gene fragments coding for full length rat HRas (AAA42009.1), human Rap1B (CAB46488), and human Rap2A (CAA31052.1) were PCR amplified and restriction cloned with EcoRI-HF and XhoI (R3101S and R0146S, NEB) into pGEX-6P-1 expression vectors in frame with an N-terminal glutathione S-transferase tag and PreScission Protease cleavage site. Sequence verified plasmids were transformed into Rosetta2 *E. coli* cells (#71402-4, Novagen) for protein expression. Single colonies were grown overnight at 37 °C in 5 ml of LB supplemented with 100 μg/ml carbenicillin and 34 μg/ml chloramphenicol. Overnight cultures were diluted 1:500 into LB supplemented with 2% glucose and grown at 37 °C for ~4-6 hours to an $O.D._{600}$ of 1.0. Cultures were then chilled to 22°C, and protein expression was induced with 0.1 mM IPTG for 4 hours. Bacterial pellets were harvested by centrifugation, resuspended in Lysis/Wash Buffer (50 mM Tris, 100 mM NaCl, 5 mM TCEP, 5 mM $MgCl_2$, 1 mM EDTA, 1mM PMSF, 30 μM GDP, 0.1% Triton X100, Complete protease inhibitor, pH 7.5) supplemented with

25 U/ml Benzonase and 10 kU/ml ReadyLyse, and lysed by fluidization. Insoluble material was removed by centrifugation at 30,000 x g. Supernatant was loaded onto Glutathione Agarose resin (#16100, Pierce) via a hybrid batch gravity procedure, washed with Lysis/Wash Buffer, and eluted using Cleavage Buffer (50 mM Tris, 100 mM NaCl, 5 mM TCEP, 5 mM $MgCl_2$, 1 mM EDTA, 30 μM GDP, 0.1% Triton X100, and 50 mM glutathione, pH 8.0). Fractions containing GTPases were pooled, and concentrated using ultrafiltration (#89885A, Pierce). Concentrated samples of GTPase were mixed with 20 U/ml PreScission Protease (27-0843-01, GE Healthcare) and cleaved off-column for 18 hours at 4°C. Samples were buffer exchanged and concentrated by ultrafiltration before being loaded onto a Hi Load 26/600 Superdex 75 prep grade column (28-9893-36, GE Healthcare) equilibrated with SEC Buffer (20 mM Tris, 100 mM NaCl, 5 mM TCEP, 5 mM $MgCl_2$, 1 mM PMSF, 30 μM GDP, and 10% Glycerol, pH 8.0). Fractions containing GTPases were pooled, concentrated by ultrafiltration to greater than 1 mg/ml, flash frozen in liquid nitrogen, and stored at -80°C.

SDS-PAGE and Assessment of Protein Purity and Yield. We used SDS-PAGE to determine purity of proteins and to quantify yields throughout the purification process. Protein samples were diluted into 4x LDS buffer (#NP0008) and heated to 70°C for 10 minutes before fractionation on NuPAGE Novex 4-12% Bis-Tris SDS-PAGE gradient gels (#NP0008) run under reducing conditions with NuPAGE Antioxidant (#NP0005) and MOPS running buffer (#NP0001), all purchased from Life Technologies. Proteins in the gels were stained with Gel Code Blue (#24592, Thermo Scientific), and imaged on

a Licor Odyssey Classic Infrared Imaging System (Licor Biosciences) at 700 nm, using Licor Image Studio Software (v3.1) for quantification. Molecular weights of stained proteins were verified by comparison to Precision Plus Protein All Blue Standards (Cat. No. 161-0373, BioRad, Hercules CA). The amount of each stained protein was measured in the gel by comparison to known amounts of bovine serum albumin (#A5611), lysozyme (#L6876), or LacZ (#G8511), all purchased from Sigma-Aldrich. The protein standards were loaded onto each gel in lanes adjacent to the protein samples.

Site Directed Mutagenesis of synGAP (S773/T775/S802 Single, Double and Triple Mutants). Ten 103-1293 synGAP mutants were prepared to study the effects of specific CDK5 phosphorylation sites on synGAP activity: S773A, T775A, S802A, S773D, T775D, and S802D single mutants; S773A/T775A, S773D/T775D, and S773D/S802D double mutants; and S773A/T775A/S802A triple mutant. Ultramer oligonucleotides coding for these specific mutations were designed using the QuikChange II Primer Design Program (Agilent) and synthesized by Integrated DNA Technologies (IDT). Mutation of synGAP was performed using the QuikChange II XL Site Directed Mutagenesis Kit (#200522, Agilent).

Site Directed Mutagenesis of synGAP (CTM-Plus Mutant). A mutant synGAP construct devoid of all previously identified CaMKII phosphorylation sites (CTM-Plus Mutant) [6] was prepared in order to isolate the effects of additional CaMKII phosphorylation sites on synGAP activity. Serine and threonine residues at sites

750/751/756/764/765/1058/1062/1064/1093/1095/ 1099/1123/1125 in synGAP were all replaced with alanine residues. A pair of complementary oligonucleotides coding for alanine mutations of 750/751/756/764/765 was designed using DNA2.0 Gene Designer (v1) (DNA2.0) and IDT Oligo Analyzer Software, and synthesized as Ultramers by IDT. Mutation of synGAP was performed with a QuikChange II XL Site Directed Mutagenesis Kit using the procedure described by [17]. A synthetic gene coding for alanine mutations of sites 1058/1062/1064/1093/1095/1099/1123/1125 in synGAP was designed as described above. The gene was inserted into the 750/751/756/764/765 alanine mutant construct using the PIPE-LIC cloning method [15, 16]. Mutated synGAP constructs were transformed into One Shot TOP10 chemically competent cells (C4040-06) for propagation and verified by DNA sequencing.

Stoichiometry and Rate of r-synGAP Phosphorylation by CaMKII or CDK5/35. Phosphorylation of synGAP by CaMKII (Purified as described in [18]) or by CDK5/p35 (#14-477M, EMD Millipore) was carried out in a reaction mixture containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 0 (CDK5) or 0.7 mM $CaCl_2$ (CaMKII), 0.4 mM EGTA, 30 or 500 μM [γ-^{32}P]-ATP (100-375 cpm/pmol) (6000 Ci/mmol, BLU002Z001MC, Perkin Elmer), 0 (CDK5) or 3.375 μM calmodulin (CaMKII), 10 mM DTT, 286 nM r-synGAP, and 3.1 nM rat brain CaMKII or 110 nM CDK5/p35. Phosphorylation was initiated by addition of ATP to a reaction mixture prewarmed to 30 °C. At the indicated time points, reaction aliquots were removed and quenched by the addition of ice-cold 4x LDS Sample Buffer containing 10% β-mercaptoethanol. Samples

were separated by SDS-PAGE as described above. Gels were rinsed in water, wrapped in Saran Wrap, and exposed to a Storage Phosphor screen (#63-0034-79, GE Healthcare) for 2-4 hours to identify phosphorylated proteins. To quantify the amount of phosphate incorporated, the density of ^{32}P bands was determined with a Typhoon LA 9000 (#28-9558-08, GE Healthcare) and ImageQuant TL software (#29000737, GE Healthcare). The relative densities measured by the Imager were converted to pmol phosphate by comparison to densities from standard amounts of [γ-^{32}P]-ATP spotted onto filter paper and imaged simultaneously. To calculate the stoichiometry of phosphorylation, the pmol of incorporated phosphate determined using ImageQuant was divided by amount of synGAP (3.8 pmol) loaded per lane.

Confirmation of synGAP Phosphorylation by CaMKII or CDK5/p35 via Western Blotting. CDK5/p35 and CaMKII catalyzed phosphorylation of synGAP were carried out as described above using 30 or 500 μM cold ATP in place of [γ-^{32}P]-ATP. Following SDS-PAGE, proteins were electrically transferred to PVDF membranes (#88520, Pierce) in 25 mM Tris, 200 mM glycine and 20% methanol. Membranes were washed with TBS, followed by blocking with Odyssey Blocking Buffer (#927-40003, Li-Cor Biosciences). Membranes were washed in TBS-T before incubation in Odyssey Blocking Buffer containing one or more of the following primary antibodies: 1:2000 mouse anti-bovine serum albumin (#ab3781, AbCaM), 1:2000 rabbit anti-phospho-synGAP-S1123 (M.B.K. laboratory), 1:2000 rabbit anti-phospho-synGAP-S764/5 (M.B.K. laboratory), 1:1000 rabbit anti-synGAP (#PA1-046, Pierce), 1:500 rabbit anti-

pSer CDK substrate (#2324, Cell Signaling), 1:500 rabbit anti-pSer CDK/MAPK substrate (#2325S, Cell Signaling) or 1:5000 anti-pThr CDK/MAPK substrate (#2321, Cell Signaling), and 1:1000 BSA-free anti-Tetra-His (#34670, Qiagen). Bound antibodies were detected with 1:10,000 goat anti-rabbit IRdye700 (#610-730-124, Rockland) or 800 (#611-132-002, Rockland) or 1:10,000 goat anti-mouse Alexa-Fluor 680 (#A-21057, Life Technologies), visualized with a Licor Odyssey Classic Infrared Imaging System, and quantified using the Licor Image Studio Software (v3.1).

Nucleotide Exchange. The nucleotide exchange procedure is based on the methods described in [19-21]. HRas or Rap (200 μM) was incubated in 25 mM Tris, pH 8.0, 100 mM NaCl, 5 mM DTT, 15 mM EDTA, 100 mM ammonium sulfate and 10 mM GTP for 60 minutes at room temperature. Exchange reactions were terminated by the addition of 25 mM $MgCl_2$. GTPases were concentrated and separated from unbound nucleotides and EDTA by ultrafiltration in an Amicon Concentrator (#UFC501096, EMD Millipore) at 4°C. Radioactive [γ-^{32}P]-GTP bound GTPase was prepared by incubating 0.8 or 1.6 mM GTP-GTPase with 10 μCi of [γ-^{32}P]-GTP (6000 Ci/mmol; #NEG004Z001MC, Perkin Elmer) in the presence of 12 mM EDTA for 40 minutes on ice. The exchange reaction was terminated with the addition of 25 mM $MgCl_2$. [γ-^{32}P]-GTP bound HRas or Rap GTPases were diluted to the desired concentration before use in GTPase assays.

GTPase Assays. synGAP stimulated hydrolysis of HRas and Rap GTPases was

assayed by measuring the release of [γ-^{32}P]-phosphate via the charcoal method [22]. Phosphorylated or non-phosphorylated synGAP (250-500 nM) was mixed with increasing amounts of radioactive [γ-^{32}P]-GTP bound GTPase at 25 °C in Assay Buffer (25 mM Tris, pH 8.0, 100 mM NaCl, 15 mM $MgCl_2$, and 5 mM DTT) as described in [20]. Initial rates of GTP hydrolysis were evaluated by linear regression, plotted against substrate concentrations, and analyzed by nonlinear regression using the Michaelis-Menten equation to calculate k_{cat} and K_M values in Prism (v6.0d, GraphPad Software, La Jolla CA).

Phosphorylation of r-synGAP by CaMKII or CDK5/35 for use in GAP Assays. Phosphorylation of synGAP by CaMKII or CDK5/p35 for use in GAP assays was carried out in reaction mixtures containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 0 (CDK5) - 0.7 mM $CaCl_2$ (CaMKII), 0.4 mM EGTA, 30 μM ATP, 0 (CDK5) - 3.375 μM calmodulin (CaMKII), 10 mM DTT, 724.5 nM r-synGAP, and 10 nM rat brain CaMKII or 230 nM CDK5/p35. Phosphorylation was initiated by addition of ATP to reaction mixtures prewarmed to 30 °C. At the indicated time points, reaction aliquots were removed and quenched by the addition of 1/3 volume of ice-cold Kinase Quench Buffer [20 mM Tris-HCl, pH 8.0, 396.3 mM NaCl, 2 mM DTT, 30 mM EGTA, 0.2% Tergitol Type NP-40, 90 μM Roscovitine (#R7772-5MG, Sigma-Aldrich) and 6 μM Autocamtide-2 related inhibitory peptide (#A4308, Sigma-Aldrich)]. Samples were stored on ice until used in GAP assays.

Mass Spectrometry of Phosphorylated r-synGAP. Mass spectrometry of phosphorylated r-synGAP was performed by the Proteome Exploration Laboratory at the California Institute of Technology. All liquid chromatography-mass spectrometry experiments were performed on an EASY-nLC (Proxeon Biosystems, now Thermo Scientific) connected to a hybrid LTQ-FT (Thermo Scientific) equipped with a nano-electrospray ion source (Proxeon Biosystems, now Thermo Scientific) essentially as previously described [23], with the modifications indicated below. Peptides were separated on a 15 cm reverse phase analytical column (75 μm ID) packed in-house with 3 μm ReproSil-Pur C18AQ beads (#r13.aq., Dr. Maisch GmbH) using a 60 minute gradient at a flow rate of 350 nl/minute. The gradient was run from 0% to 40% solvent B (97.8% acetonitrile/0.2% formic acid/2% water) in 40 minutes, increased to 100% B in 1 minute, and run at 100% B for 19 minutes. The mass spectrometer was operated in data-dependent mode to automatically switch between full-scan MS and tandem MS acquisition. Survey full scan mass spectra were acquired in FT (400–1800 m/z), following accumulation of 1,000,000 ions, with a resolution of 100,000 at 400 m/z. The top five most intense ions from the survey scan were isolated and, after the accumulation of 5000 ions, fragmented in the linear ion trap by collision-induced dissociation (collisional energy 35% and isolation width 2 Da). Additionally, an MS3 neutral loss scan was acquired if a neutral loss of 32.67, 49.00, or 97.97 Da was observed in the top five most intense peaks in the MS2. Precursor ion charge state screening was enabled, and all singly charged and unassigned charge states were rejected. The dynamic exclusion list was set with a maximum retention time of 90 seconds and a relative mass

window of 10 ppm, and early expiration was enabled.

Tandem mass spectra were converted to mgf files using ReAdW4Mascot2 (http://peptide.nist.gov/metrics/). Additionally, MS3 spectra were separated from MS2 spectra and accurate masses determined for MS3 precursor spectra following Timm et al. [24]. All MS/MS samples were analyzed using Mascot (v.2.2.06, MatrixScience). A sequence database was constructed, consisting of the sequence of synGAP and common contaminants, including keratins and trypsin (262 entries). Trypsin was specified as the digestion enzyme, and up to two missed cleavages were allowed. Mass tolerances were set to 10 ppm for parent ions and 0.50 Da for fragment ions. Carbamidomethylation of cysteine (+57.0215 Da) was specified in Mascot as a fixed modification, and oxidation of methionine (+15.9949 Da), dehydration of serine and threonine (-18.0106 Da, only MS3 spectra), and phosphorylation of serine, threonine, and tyrosine (+79.9663 Da, only MS2 spectra) were specified as variable modifications. Mascot results were loaded in Scaffold (v 3.00.06, Proteome Software) and filtered at a protein probability of 99%, with a minimum of two peptides and a peptide probability of 95%. Phosphorylation site false location rates were calculated using the MD-score, as described by Savitski *et al.* [25]. Raw data and annotated spectra for all phosphorylation sites will be uploaded and made publicly available following manuscript acceptance.

RESULTS AND DISCUSSION

Identification of a CaMKII-Independent Phosphorylation Site in synGAP. To identify novel sites of CaMKII-independent phosphorylation on synGAP, we subjected the PSD fraction to phosphorylation in the presence of Mg^{2+}/ATP, but the absence of Ca^{2+} and CaM, as described in Methods. After phosphorylation, proteins were fractionated by SDS-PAGE, and the 135 kDa band containing synGAP was excised for tryptic digestion. Liquid chromatography and tandem mass spectrometric analysis of the digest revealed a phosphopeptide corresponding to the peptide containing S773 in synGAP (Table 3.1). Figure 3.1 shows the ion fragmentation spectrum from LC/MS/MS analysis of the doubly charged phosphopeptide with an observed *m/z* of 361.6906 Da. The b and y ion fragmentation series that we used to determine the sequence, LPSpPTK, are marked. Based on the phosphorylation conditions and on the sequence surrounding the phosphorylated serine, we reasoned that it is unlikely that CaMKII phosphorylates this site. Instead, the sequence surrounding the site closely resembles the consensus sequence for the proline directed kinase, CDK5 (SPXR/K). CDK5 has not previously been shown to phosphorylate or regulate synGAP.

Purification of r-synGAP. It has previously been shown that residues 103-725 of synGAP can be expressed solubly in *E. coli* and form a stable protein whose catalytic activity is similar to other RapGAPs, but well below that of other RasGAPs [4]. While this soluble construct contains synGAP's N-terminal pleckstrin homology (PH) domain, C2 domain, and RasGAP domain, it lacks the disordered domain that contains regulatory

CaMKII phosphorylation sites [6] and the C-terminal PDZ domain binding motif [1, 2]. In order to facilitate a study of the regulatory roles of CaMKII and CDK5 phosphorylation on synGAP, we expressed and purified a nearly full length, soluble 103-1293 synGAP fragment (r-synGAP) (Fig. 3.2A). To assess additional affects of the regulatory role of the disordered region on basal r-synGAP activity, we also purified the 103-725 r-synGAP fragment for comparison (Fig. 3.2B). Both r-synGAP fragments were soluble at low expression temperatures, but required ATP and magnesium to release bound chaperone (GroEL) for effective binding to Talon Co^{2+}-CMA columns (Data not shown). Using a combination of immobilized metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) , as described in Methods, 103-725 r-synGAP was purified to >95% purity, with typical yields in excess of 75%. 103-1293 r-synGAP was purified using IMAC to ~70% purity, with yields in excess of 85%. Interestingly, 103-1293 r-synGAP eluted in the void volume of SEC columns (No improvement of protein purity), suggesting that addition of the disordered domain may predispose r-synGAP to aggregation.

Purification of HRas, Rap1AB and Rap2A. Purification of Ras and Rap has been previously performed [1, 20, 26, 27]. Because our exact purification procedure (described in Methods) has not been previously published, we include it in this chapter. Ras and Rap were expressed in *E. coli* as GST fusion proteins and purified using glutathione agarose (Fig. 3.3). To remove the N-terminal GST moiety, purified GST-GTPases were cleaved with PreScission protease. Size exclusion chromatography (SEC)

was used to separate uncleaved GST-GTPase from free GST and free GTPase. Using the described purification procedure, all three GTPases were purified to >95% purity, with yields typically on the order of 25%.

Intrinsic GAP Activity of r-rsynGAP. In spite of synGAP's homology to the RasGAPs, p120GAP, and neurofibromin, synGAP stimulates the GTPase activity of Rap much more potently than of Ras [3, 4]. The 103-725 fragment of r-synGAP containing its C2 and RasGAP domains was previously shown to stimulate the intrinsic Rap1 GAP activity by four orders of magnitude, with a k_{cat} of 0.8 s^{-1}, and stimulate the intrinsic HRas GAP activity by two orders of magnitude [4]. The RapGAP activity of synGAP was also shown to be unique in that it required the presence of its C2 domain, which makes synGAP the only example of a RapGAP that utilizes a second domain for its catalytic activity, and suggested a new function of C2 domains.

We attempted to determine the kinetic profiles and constants for the 103-725 r-synGAP and 103-1293 r-synGAP constructs for the HRas, Rap1, and Rap2 substrates (Fig. 3.4). Under the majority of substrate concentrations tested, the rate of r-synGAP activity was linear, and thus precluded accurate calculation of k_{cat} and K_M.

Under the range of HRas substrate concentrations tested (10-700 μM), both 103-725 and 103-1293 r-synGAP exhibited linear rates of HRas hydrolysis, making calculated values of k_{cat} and K_M ambiguous (Fig. 3.4A). Both r-synGAP constructs can stimulate the rate of GTP hydrolysis by two orders of magnitude, which is consistent with previously published data [4].

In contrast to data obtained for HRas, 103-725 r-synGAP's Rap1 GAP activity followed a hyperbolic reaction profile allowing calculation of k_{cat} (0.5 ± 0.06 s^{-1}) and K_M (203 ± 55 μM) (Fig. 3.4B). The calculated k_{cat} value for Rap1 was in good agreement with the value calculated in [4] (k_{cat} of 0.8 s^{-1}). 103-1293 r-synGAP catalyzed the inactivation of Rap1 more rapidly than 103-725 and had a higher catalytic constant (k_{cat} of 3.8 ± 1.1 s^{-1}) (Fig. 3.4B). Interestingly, the affinity of 103-1293 r-synGAP for Rap1 was much lower (K_M of 1819 ± 600 μM) than 103-725 r-synGAP (K_M of 203 ± 55 μM). The k_{cat}'s of 103-725 and 103-1293 r-synGAP were 4,000- and 30,000-fold greater, respectively, relative than Rap1's intrinsic rate of GTP hydrolysis (0.00013 ± 0.00008 s^{-1}). The disordered domain and additional regulatory regions present in 103-1293 r-synGAP appear to help stimulate the rate of Rap1 hydrolysis approximately 8-fold over 103-725 r-synGAP; however, they also seem to substantially lower the affinity of r-synGAP for Rap1.

Under the range of Rap2 substrate concentrations tested (15-400 μM), 103-725 r-synGAP's GAP activity was linear, precluding accurate calculation of k_{cat} and K_M (Fig. 3.4C). 103-1293 r-synGAP catalyzed the inactivation of Rap2 at approximately the same rate as 103-725 (Fig. 3.4C). Its reaction profile was slightly hyperbolic which allowed a rough estimation of its catalytic constant (k_{cat} of 5.5 ± 2.1 s^{-1}) and K_M (2466 ± 1056 μM). The k_{cat} of 103-1293 r-synGAP was approximately 70,000-fold greater than Rap2's intrinsic rate of GTP hydrolysis (0.00008 ± 0.000002 s^{-1}). The effect of the disordered domain and additional regulatory regions present in 103-1293 r-synGAP on the catalysis of Rap2 cannot be assessed, due to the ambiguity of kinetic constants

calculated for the Rap2 and 103-725 r-synGAP reaction.

Effect of CaMKII Phosphorylation on r-synGAP Activity. We have previously reported that the Ras GAP activity of synGAP is activated by 75% in PSD bound synGAP and by 95% in Sf9 membrane bound r-synGAP [6]. To determine if CaMKII phosphorylation affects r-synGAP's activity, we measured the GAP activity of r-synGAP before and after CaMKII phosphorylation, as described in Methods (Fig. 3.5). Phosphorylation by CaMKII increased r-synGAP's HRas and Rap1 GAP activity, but not its Rap2 GAP activity. In the absence of phosphorylation, r-synGAP stimulated the GAP activity of 10 μM HRas and Rap (both Rap1 and Rap2) by approximately 3-fold and 11-fold, respectively. CaMKII phosphorylation of r-synGAP further increased RasGAP activity by up to 25% and Rap1GAP activity by up to 76%. CaMKII phosphorylation of r-synGAP produced a steady increase in RasGAP activity that plateaued at 10 minutes, mirroring the reaction profile of CaMKII phosphorylation (Discussed below). Interestingly, Rap1GAP activity was increased almost immediately upon incubation with CaMKII. After 1 minute of CaMKII phosphorylation, Rap1GAP activity increased by 61%, and after 2 minutes it reached a maximal activity increase of approximately 75%. CaMKII phosphorylation of r-synGAP resulted in an insignificant change in Rap2GAP activity. Thus, phosphorylation of synGAP by CaMKII accelerated the rate of inactivation of Rap1 more potently than of HRas, and may lead to alteration of the ratio of activated Ras and Rap in synapses.

Effect of CDK5 Phosphorylation on r-synGAP Activity. To determine if CDK5 phosphorylation affects r-synGAP's activity, we measured the GAP activity of r-synGAP before and after CDK5 phosphorylation, as described in Methods (Fig. 3.6). Phosphorylation by CDK5 significantly increased r-synGAP's HRas and Rap1 GAP activity, but not its Rap2 GAP activity. CDK5 phosphorylation of r-synGAP increased RasGAP activity by up to 65% and Rap1GAP activity by up to 20%. CDK5 phosphorylation of r-synGAP produced a steady increase in RasGAP activity that plateaued at 5 minutes. Similar to results obtained with CaMKII, Rap1GAP activity increased almost immediately upon incubation with CDK5. After 1 minute of CDK5 phosphorylation, Rap1GAP activity increased by 15%, and after 5 minutes it reached a maximal activity increase of 19%. Like CaMKII, CDK5 phosphorylation of r-synGAP did not produce a statistically significant change in Rap2GAP activity. Thus, phosphorylation of synGAP by CDK5 accelerated the rate of inactivation of HRas more potently than of Rap1 (The opposite effect of CaMKII phosphorylation), and may lead to alteration of the ratio of activated Ras and Rap in synapses.

Stoichiometry and Rate of r-synGAP Phosphorylation by CaMKII. r-synGAP is phosphorylated rapidly and to a high stoichiometry by CaMKII (Fig. 3.7A). The most rapid phosphorylation occurs within 2 minutes, after which r-synGAP contains ~6 mol of phosphate/mol r-synGAP, suggesting that there are six preferred sites of phosphorylation by CaMKII. After 10 minutes the rate essentially reaches an asymptote at a stoichiometry of approximately 9 mol phosphate incorporated per pmol r-synGAP.

Because our GAP assays were inhibited by concentrations of ATP in excess of 30 μM, we compared the stoichiometry of r-synGAP phosphorylation by CaMKII at 30 and 500 μM ATP. Reaction mixtures containing 30 μM ATP followed the same reaction profile as reactions with 500 μM ATP; however, the overall stoichiometry plateaued at approximately 6.5 mol phosphate incorporated per pmol r-synGAP (data not shown). The stoichiometry and reaction profile of r-synGAP phosphorylation by CaMKII at 30 μM ATP were nearly identical to phosphorylation reactions utilizing endogenous synGAP and CaMKII in the PSD [6].

Using previously developed phosphosite-specific antibodies against phosphorylated peptides containing S765 and S1123 [6], we tracked the site occupancy of S765 and S1123 in CaMKII phosphorylated r-synGAP by immunoblotting (as described in Methods) (Fig. 3.7B). While both S765 and S1123 were deemed major sites of CaMKII phosphorylation in a previous paper [6], we have determined that their reaction profiles differ significantly. S1123 occupancy roughly followed the overall stoichiometry of phosphorylation (Fig. 3.7A), reaching approximately 80% occupancy in 2 minutes, and hitting a plateau at 10 minutes. S765 was phosphorylated at much slower rate, with the percent coverage increasing linearly over the 10 minute reaction period. No additional phosphorylation occurred at S765 or S1123 after 10 minutes (data not shown). Therefore, the occupancy of S765 and S1123 sites at specific time points was normalized to the 10 minute time point.

Stoichiometry and Rate of r-synGAP Phosphorylation by CDK5. With CDK5

the most rapid phosphorylation of r-synGAP occurred within 10 minutes, after which r-synGAP contained ~0.3 to 0.4 mol phosphate/mol r-synGAP (Fig. 3.8). After approximately 10 minutes, the rate slowed. Overall, the curve approximated a hyperbolic shape, with an asymptote approaching a stoichoimetry of 1 pmol phosphate/pmol r-synGAP. These data suggested that there is a single preferred site of phosphorylation by CDK5. r-synGAP was phosphorylated relatively slowly and to a lower stoichiometry by CDK5 (Fig. 3.8) than by CaMKII (Fig. 3.7A). These differences may reflect differences in the intrinsic properties of the phosphorylating enzymes (CaMKII versus CDK5) or differences in the respective reaction mixtures. The reaction mixtures for CaMKII contained a 100:1 molar ratio of r-synGAP (substrate) to kinase, while those for CDK5 contained a 2:1 ratio molar ratio of r-synGAP to kinase. Phosphorylation of 300 nM synGAP by CDK5/p35 occurred at a similar rate (0.03 mol phosphate/mol substrate*min) to phosphorylation of the CDK5 substrate histone H1 (4.3 μM; data not shown), indicating that r-synGAP is a comparable substrate to histone H1.

Unfortunately, all attempts to track CDK5 phosphorylation at S773 via immunoblotting were unsuccessful (data not shown). Commercially available phosphosite-specific antibodies developed against phosphorylated peptides containing CDK5 consensus sequences (PXS*P; S*PXR/K, where * indicates phosphorylation of the immediately preceding amino acid) or mitogen activated protein kinase (MAPK) consensus sequences (T*P) (as described in Methods) did not produce any detectable signal with immunoblotting. Mass spectrometry (described below) was able to confirm CDK5 phosphorylation at S773 and S802 consensus sequences in synGAP.

Identification of CaMKII Phosphorylation Sites in r-synGAP. We have previously shown by mass spectrometry that membrane bound r-synGAP is phosphorylated by CaMKII at multiple sites, including S750/S751/S756, S764/S765, S1058, S1093/1095, S1099, and S1123 [6]. Using mass spectrometry (as described in Methods), we were able to confirm that previously identified CaMKII phosphorylation sites are phosphorylated in 103-1293 r-synGAP (S751, S765, S1093, S1095, S1099, and S1123) (Table 3.2). We were also able to identify several novel CaMKII phosphorylation sites in r-synGAP (S737, S821, S825, S827, S882, T897, S1150, S1210, and S1283) (Table 3.2). CaMKII phosphorylation reactions with r-synGAP were carried out for 2 or 10 minutes and quenched with LDS stop buffer, as described in Methods. Following separation by SDS-PAGE, phosphorylated r-synGAP was reduced, alkylated, and in-gel digested with trypsin before separation by nano-LC and analysis on a hybrid mass spectrometer equipped with a linear ion trap in combination with Fourier transform ion cyclotron resonance. Ions collected in the linear ion trap were fragmented by collision induced dissociation (MS2). If a neutral loss of 32.67, 49.00, or 97.97 Da was observed in MS2, an MS3 neutral loss scan was acquired to detect the presence of dehydroalanine at phosphorylation sites, resulting from the β-elimination of H_3PO_4 from phosphoserine. Following completion of data collection and analysis via Mascot, results were stringently filtered to a protein probability of 99%, a two peptide minimum, and peptide probability of 95% in Scaffold (as described in Methods). The false localization rate of phosphorylation site identifications was then assessed using the Mascot Delta

score as described in [25]. Peptide sequence coverage of r-synGAP 103-1293 was in excess of 84%, with 90% coverage of the disordered domain (91.9% theoretical maximum coverage of 103-1293 synGAP calculated by ExPASy Peptide Mass Program). CaMKII phosphorylation sites identified using this workflow are shown in Table 3.2. As stated above, the previously identified CaMKII sites S751, S765, S1093, S1095, S1099, and S1123 were phosphorylated in 103-1293 r-synGAP. While we did detect phosphorylation at sites S750, S756, and S764 (primarily at 10 minute time points), calculated phosphorylation false localization rates at these sites were >5%. Since sites S751, S765, S1099, and S1123 had false localization rates under 1%, we believe that these four serines are the primary sites of CaMKII phosphorylation in r-synGAP. Previous work by Oh *et al.* [6] suggested that S1093/S1095 and S1058 were minor sites of CaMKII phosphorylation in membrane bound r-synGAP. Our results suggest that both S1093 and S1095 are primary sites of phosphorylation in r-synGAP; however, S1058 was not detected in our experiments. In our mass spectrometry experiments we also identified several novel sites of phosphorylation in r-synGAP including, S737, S821, S825, S827, S882, S883, T897, S1150, S1210, and S1283. All of these sites were detected at both the 2 minute and 10 minute time points and had false localization rates of less than 1%, suggesting that they are high affinity CaMKII phosphorylation sites. We detected a large number phosphorylation sites in residues 800-900 of r-synGAP, suggesting that this region may be surface exposed or particularly disordered. Interestingly, CaMKII appears to phosphorylate a residue, S1283, in close proximity to r-synGAP's C-terminal PDZ domain binding ligand, QTRV (1290-1293).

Identification of CDK5 Phosphorylation Sites in r-synGAP. In Figure 3.1 we used mass spectrometry to identify a CaMKII-independent phosphorylation site (S773) in PSD bound synGAP. Using mass spectrometry (as described in Methods), we were able to confirm that this CaMKII-indepenent phosphorylation site was phosphorylated in 103-1293 r-synGAP, along with several additional sites (S728, T775, S802, and S842) (Table 3.3). CDK5 phosphorylation of r-synGAP and mass spectrometry were carried out as described above for CaMKII and in Methods. Of the identified CDK5 phosphorylation sites in r-synGAP, only S773 and S802 are located in CDK5 consensus sequences (S/TP motifs). While S773 exhibited a phosphorylation false localization rate of less than 1%, S802 had a rate greater than 5%. Close inspection of mass spectra from peptides predicted to be phosphorylated at S802, and high mascot scores (~65-85) in their MS2 and MS3 spectra confirmed the presence of phosphorylation at S802. Because of its close proximity to S773, phosphorylation at T775 may serve an analagous function, or it may result from promiscuity in the phosphorylation reaction. Because of its proximity to S773, we treated T775 as a potential CDK5 phosphorylation site. S728 and S842 may be legitimate CDK5 phosphorylation sites in r-synGAP (1093 was detected only in MS3); however, we chose to focus our efforts on S773/T775 and S802 because of their localization to CDK5 consensus sequences.

Intrinsic GAP Activity of r-synGAP Mutants. To determine the effects of CDK5 phosphorylation at sites S773 and S802 in r-synGAP, we generated a series of

103-1293 r-synGAP mutants in which several individual phosphorylation sites or combinations thereof were mutated to alanine (Phosphodeficient) or aspartic acid (Phosphomimetic). Mutants were purified with the same procedure used for wild type 103-1293 r-synGAP, as described in Methods (Fig. 3.9). We measured the Ras and Rap GAP activity of these r-synGAP mutants, as described in Methods (Fig. 3.10). As expected, mutation of T775A and S802A did not affect GAP activity. Surprisingly, the S773A single mutant and S773A/T775A double mutant had significantly reduced GAP activity. This is in contrast to the S773A/T775A/S802A triple mutant that exhibited increased Ras and Rap2 activity, all of which suggest that mutation of S773 may disrupt proper folding of synGAP or affect catalysis in a non-trivial way. The CTM-Plus mutant of r-synGAP was strongly inhibited when tested with Ras or Rap, suggesting that bulk mutagenesis of disordered domain residues to alanine affects protein folding. The GAP activity of single and double aspartic acid mutants S773D, T775D, and S773D/T775D was strongly inhibited, suggesting that S773 or T775 phosphorylation may serve an inhibitory role. In contrast, the S802D single mutant and S773D/S802D double mutant showed elevated GAP activity, suggesting that S802 phosphorylation may have an activating role. While we expected changes in HRas and Rap1 GAP activity with phosphomimetic S773D, T775D, and S802D mutations, we did not expect changes in Rap2 GAP activity. This result is perplexing, and suggests that data generated using phosphomimetic r-synGAP mutants may not in all cases accurately represent the effects of phosphorylation. Additionally, while the general trend of activation or inhibition correlated with specific mutations consistently across each tested GTPase, the degree of

activation or inhibition could be markedly different. Inhibitory phosphomimetic mutations (S773D, T775D, etc.) did not inhibit Rap1 GAP activity nearly as potently as HRas or Rap2 GAP activity. To further assess the affects of CDK5 phosphorylation at specific sites in r-synGAP, we will phosphorylate the phosphomimetic and phosphodeficient mutants with CDK5 and examine the resulting changes in activity. If our hypothesis is correct, phosphorylation of the S773D or T775D r-synGAP mutants with CDK5 should only result in phosphorylation at S802, and would be expected to result in increased synGAP activity. Phosphorylation of S802D r-synGAP mutants with CDK5 should result in phosphorylation at S773 or T775, and would be expected to result in decreased GAP activity.

ADDITIONAL EXPERIMENTS TO COMPLETE BEFORE MANUSCRIPT SUBMISSION

Repeat of k_{cat} and K_M Determinations for r-synGAP 103-725 and 103-1293. We would like to reduce the error in our current estimations of Michaelis-Menten constants, k_{cat} and K_M, for r-synGAP with HRas-, Rap1- and Rap2-GTP as substrates. The most likely source of error in the data resulted from the use of substrate concentrations well below their K_M values. We will repeat these experiments with increased concentrations of substrate.

Effect of CDK5 and CaMKII Phosphorylation on Mutant synGAP Activity. Mutation of serine and threonine residues in r-synGAP to aspartic acid has allowed us to hypothesize the effects of CDK5 phosphoryation at S773/T775 (inhibitory) and S802 (activating). If our hypothesis is correct, phosphorylation of the S773D or T775D r-synGAP mutants with CDK5 should only result in phosphorylation at S802, and would be expected to result in increased synGAP activity. Phosphorylation of S802D r-synGAP mutants with CDK5 should result in phosphorylation at S773 or T775, and would be expected to result in decreased GAP activity.

REFERENCES

[1] H.J. Chen, M. Rojas-Soto, A. Oguni, M.B. Kennedy, A synaptic Ras-GTPase activating protein (p135 SynGAP) inhibited by CaM kinase II. Neuron 20 (1998) 895-904.

[2] J.H. Kim, D. Liao, L.F. Lau, R.L. Huganir, SynGAP: a synaptic RasGAP that associates with the PSD-95/SAP90 protein family. Neuron 20 (1998) 683-691.

[3] G. Krapivinsky, I. Medina, L. Krapivinsky, S. Gapon, D.E. Clapham, SynGAP-MUPP1-CaMKII synaptic complexes regulate p38 MAP kinase activity and NMDA receptor-dependent synaptic AMPA receptor potentiation. Neuron 43 (2004) 563-574.

[4] V. Pena, M. Hothorn, A. Eberth, N. Kaschau, A. Parret, L. Gremer, F. Bonneau, M.R. Ahmadian, K. Scheffzek, The C2 domain of SynGAP is essential for stimulation of the Rap GTPase reaction. EMBO reports 9 (2008) 350-355.

[5] J.J. Zhu, Y. Qin, M. Zhao, L. Van Aelst, R. Malinow, Ras and Rap control AMPA receptor trafficking during synaptic plasticity. Cell 110 (2002) 443-455.

[6] J.S. Oh, P. Manzerra, M.B. Kennedy, Regulation of the neuron-specific Ras GTPase-activating protein, synGAP, by Ca2+/calmodulin-dependent protein kinase II. The Journal of biological chemistry 279 (2004) 17980-17988.

[7] M. Angelo, F. Plattner, K.P. Giese, Cyclin-dependent kinase 5 in synaptic plasticity, learning and memory. Journal of neurochemistry 99 (2006) 353-370.

[8] Z. Fu, S.H. Lee, A. Simonetta, J. Hansen, M. Sheng, D.T. Pak, Differential roles of Rap1 and Rap2 small GTPases in neurite retraction and synapse elimination in hippocampal spiny neurons. Journal of neurochemistry 100 (2007) 118-131.

[9] B.S. Li, M.K. Sun, L. Zhang, S. Takahashi, W. Ma, L. Vinade, A.B. Kulkarni, R.O. Brady, H.C. Pant, Regulation of NMDA receptors by cyclin-dependent kinase-5. Proceedings of the National Academy of Sciences of the United States of America 98 (2001) 12742-12747.

[10] M.A. Morabito, M. Sheng, L.H. Tsai, Cyclin-dependent kinase 5 phosphorylates the N-terminal domain of the postsynaptic density protein PSD-95 in neurons. The Journal of neuroscience : the official journal of the Society for Neuroscience 24 (2004) 865-876.

[11] R. Dhavan, P.L. Greer, M.A. Morabito, L.R. Orlando, L.H. Tsai, The cyclin-dependent kinase 5 activators p35 and p39 interact with the alpha-subunit of Ca2+/calmodulin-dependent protein kinase II and alpha-actinin-1 in a calcium-dependent manner. The Journal of neuroscience : the official journal of the Society for Neuroscience 22 (2002) 7879-7891.

[12] K.O. Cho, C.A. Hunt, M.B. Kennedy, The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. Neuron 9 (1992) 929-942.

[13] R.K. Carlin, D.J. Grab, R.S. Cohen, P. Siekevitz, Isolation and characterization of postsynaptic densities from various brain regions: enrichment of different types of postsynaptic densities. The Journal of cell biology 86 (1980) 831-845.

[14] A. Shevchenko, H. Tomas, J. Havlis, J.V. Olsen, M. Mann, In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nature protocols

1 (2006) 2856-2860.

[15] H.E. Klock, E.J. Koesema, M.W. Knuth, S.A. Lesley, Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins 71 (2008) 982-994.

[16] H.E. Klock, S.A. Lesley, The Polymerase Incomplete Primer Extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. Methods in molecular biology (Clifton, NJ) 498 (2009) 91-103.

[17] W. Wang, B.A. Malcolm, Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. BioTechniques 26 (1999) 680-682.

[18] S.G. Miller, M.B. Kennedy, Distinct forebrain and cerebellar isozymes of type II Ca2+/calmodulin-dependent protein kinase associate differently with the postsynaptic density fraction. The Journal of biological chemistry 260 (1985) 9039-9046.

[19] S. Kupzig, D. Deaconescu, D. Bouyoucef, S.A. Walker, Q. Liu, C.L. Polte, O. Daumke, T. Ishizaki, P.J. Lockyer, A. Wittinghofer, P.J. Cullen, GAP1 family members constitute bifunctional Ras and Rap GTPase-activating proteins. The Journal of biological chemistry 281 (2006) 9891-9900.

[20] T. Brinkmann, O. Daumke, U. Herbrand, D. Kuhlmann, P. Stege, M.R. Ahmadian, A. Wittinghofer, Rap-specific GTPase activating protein follows an alternative mechanism. The Journal of biological chemistry 277 (2002) 12525-12531.

[21] G. Bollag, F. McCormick, Intrinsic and GTPase-activating protein-stimulated Ras GTPase assays. Methods in enzymology 255 (1995) 161-170.

[22] C.M. Leupold, R.S. Goody, A. Wittinghofer, Stereochemistry of the elongation factor Tu X GTP complex. European journal of biochemistry / FEBS 135 (1983) 237-241.

[23] A. Kalli, S. Hess, Effect of mass spectrometric parameters on peptide and protein

identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer. Proteomics 12 (2012) 21-31.

[24] W. Timm, N. Ozlu, J.J. Steen, H. Steen, Effect of high-accuracy precursor masses on phosphopeptide identification from MS3 spectra. Analytical chemistry 82 (2010) 3977-3980.

[25] M.M. Savitski, S. Lemeer, M. Boesche, M. Lang, T. Mathieson, M. Bantscheff, B. Kuster, Confident phosphorylation site localization using the Mascot Delta Score. Molecular & cellular proteomics : MCP 10 (2011) M110 003830.

[26] J. John, M. Frech, A. Wittinghofer, Biochemical properties of Ha-ras encoded p21 mutants and mechanism of the autophosphorylation reaction. The Journal of biological chemistry 263 (1988) 11792-11799.

[27] J. Tucker, G. Sczakiel, J. Feuerstein, J. John, R.S. Goody, A. Wittinghofer, Expression of p21 proteins in Escherichia coli and stereochemistry of the nucleotide-binding site. The EMBO journal 5 (1986) 1351-1358.

FIGURES

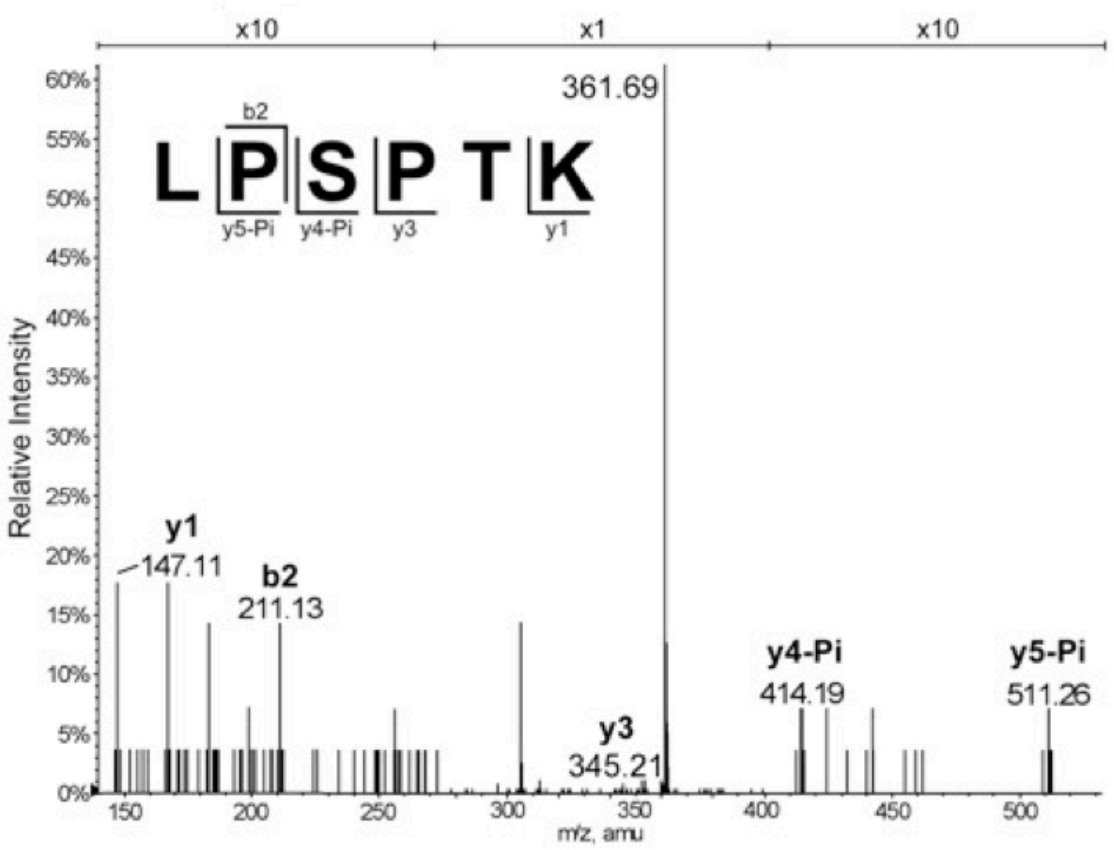


Figure 3.1. Identification of Phosphorylation of synGAP at Serine 773 by Mass Spectrometry. The doubly charged ion m/z 361.69 derived from phosphorylated synGAP isolated from PSD fraction was subjected to MS/MS analysis, as described in Methods. The observed fragment ions and the corresponding b and y ion series are indicated, showing neutral loss of phosphoric acid (-98 Da) and identifying the peptide as LPSpPTK. The peaks for m/z 140-520 are plotted as relative intensity of base peak at m/z 86.

Figure 3.2. Purification of Heterologously Expressed r-synGAP. Heterologously expressed 103-725 and 103-1293 r-synGAP proteins were purified on Talon metal affinity resin and Superdex 200 size exclusion resin as described in Methods: (**A**) Purification of His-tagged r-synGAP 103-1293; and (**B**) Purification of His-tagged r-synGAP 103-725. Fractions from each purification and Precision Plus Protein All Blue Standards were separated on a 4-12% gradient SDS-PAGE gel and stained with Gel Code Blue. T, total protein; S, soluble protein; W, wash; E, Talon Co^{2+}-CMA eluate; X, Superdex 200 eluate.

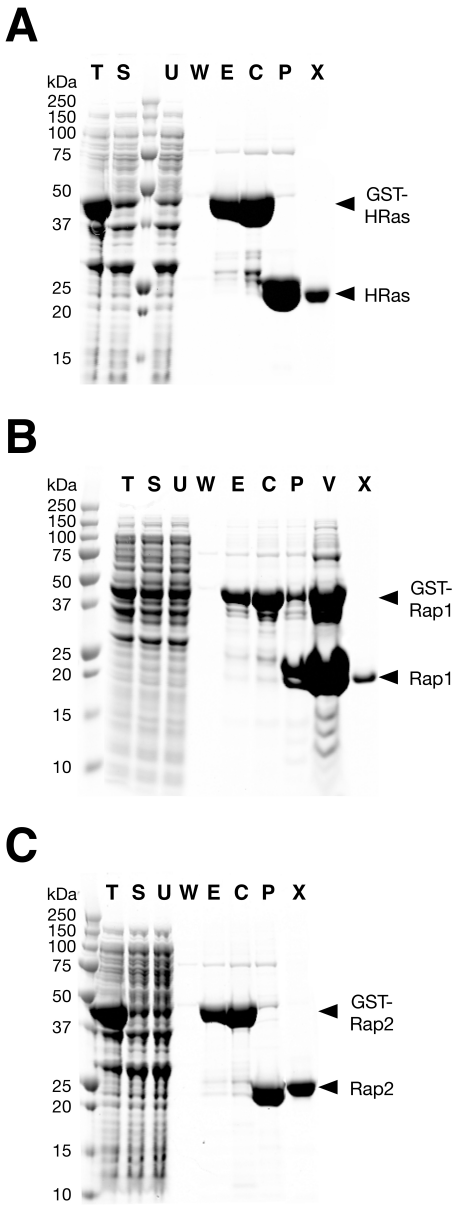


Figure 3.3. Purification of Heterologously Expressed HRas, Rap1 and Rap2. Heterologously expressed HRas, Rap1, and Rap2 proteins were purified on glutathione agarose affinity resin, cleaved off-column with PreScission protease, and separated from free GST by Superdex 75 size exclusion resin as described in Methods. Purification of (**A**) HRas; (**B**) Rap1; and (**C**) Rap2. Fractions from each purification and Precision Plus Protein All Blue Standards were separated on a 4-12% gradient SDS-PAGE gel and stained with Gel Code Blue. T, total protein; S, soluble protein; U, unbound protein; W, wash; E, glutathione agarose eluate; C, concentrated glutathione agarose eluate; P, PreScission protease cleaved; V, concentrated PreScission protease cleaved; X, Superdex 75 eluate.

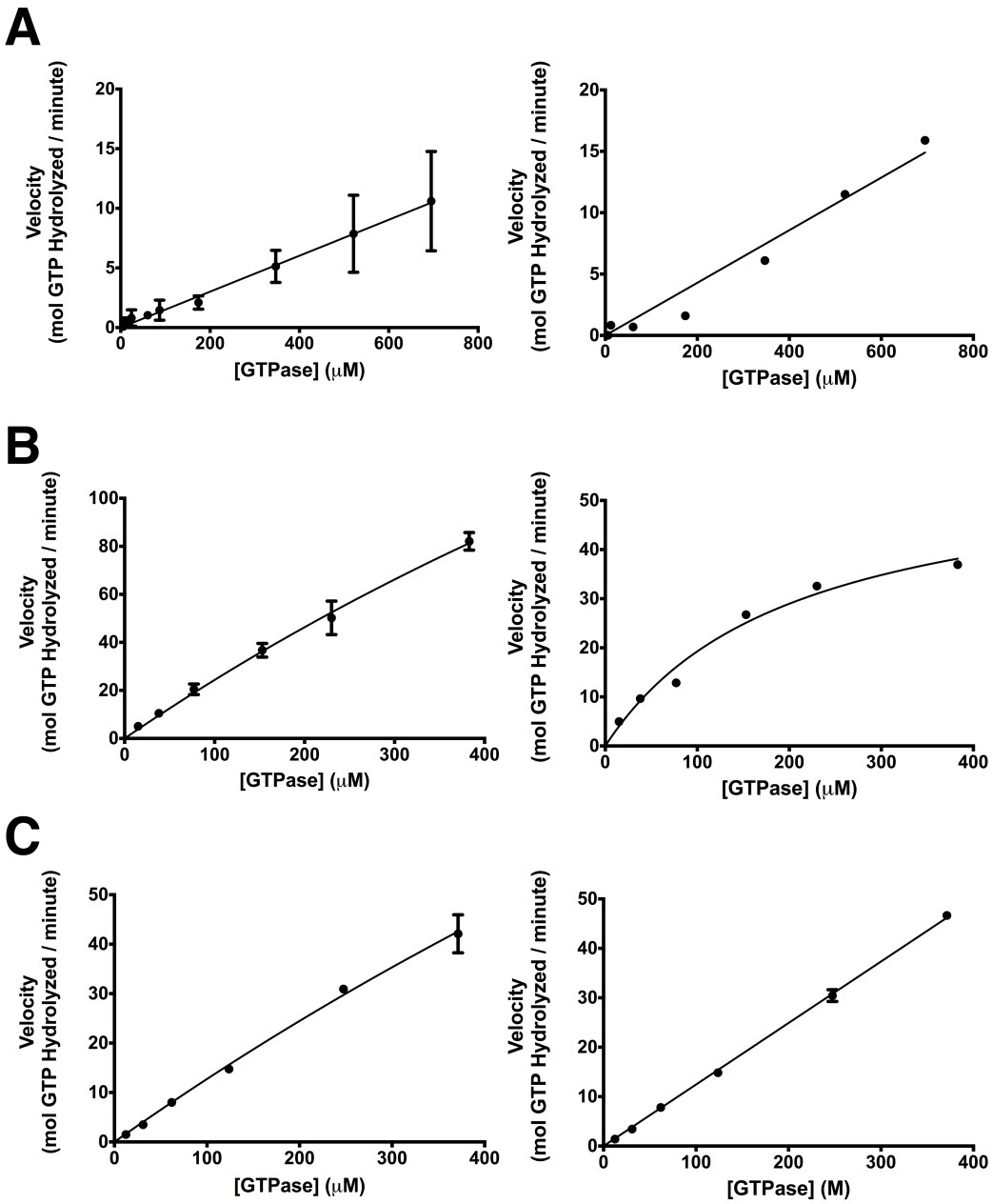


Figure 3.4. Intrinsic GAP Activity of r-synGAP. Michaelis-Menten plots for r-synGAP 103-725 (left panels) and 103-1293 (right panels) for (**A**) HRas; (**B**) Rap1; and (**C**) Rap2 substrates. GAP activity of purified r-synGAP (350 nM) was assayed with (**A**) 10-700 μM HRas- ; (**B**) 15-400 μM Rap1- ; and (**C**) 15-400 μM Rap2-[γ-^{32}P]GTP for 10 minutes at 25 °C. Michaelis-Menten constants were calculated as described in Methods. Data are mean values (n = 3-6 for 103-1293, n = 1-2 for 103-725); error bars are standard errors.

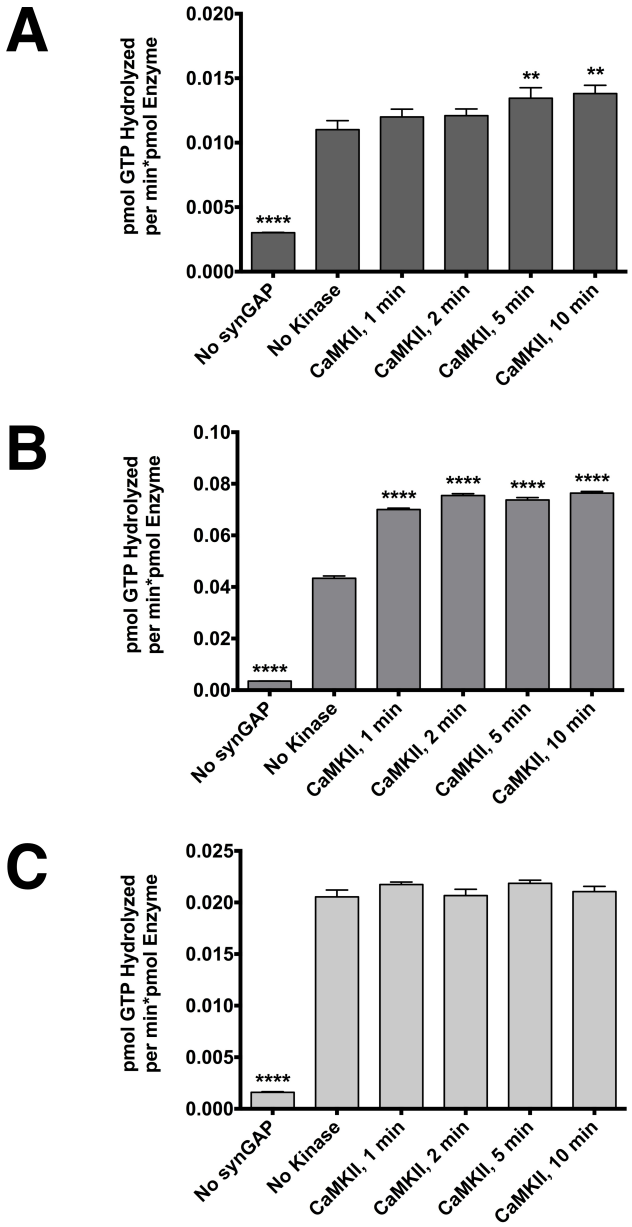


Figure 3.5. Effect of CaMKII Phosphorylation on GAP Activity of r-synGAP. GAP activity of purified r-synGAP (250 nM) before and after phosphorylation by CaMKII for 1 to 10 minutes was assayed after incubation with 10 μM (**A**) HRas- ; (**B**) Rap1- ; and (**C**) Rap2-[ɣ-^{32}P]GTP for 10 minutes at 25 °C. Initial rates (linear for 25 minutes; data not shown) were determined at single time points (10 minutes) and plotted in a bar diagram. Data are mean values (n = 9); error bars are standard errors. Significance was determined by a ordinary one way ANOVA (Uncorrected Fisher's LSD). Mean differs significantly from No Kinase: * $p<0.05$, ** $p<0.01$, *** $p<0.005$, **** $p<0.001$.

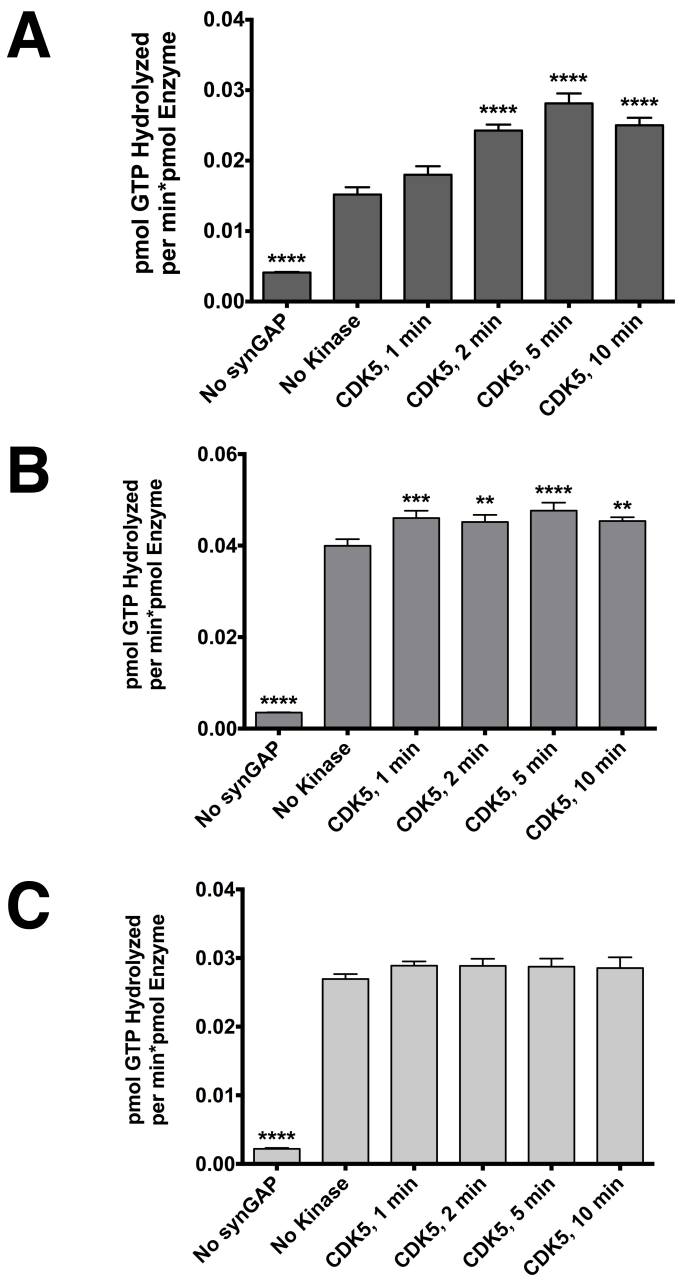


Figure 3.6. Effect of CDK5 Phosphorylation on GAP Activity of r-synGAP. GAP activity of purified r-synGAP (250 nM) before and after phosphorylation by CDK5 for 1 to 10 minutes was assayed after incubation with 10 μM (**A**) HRas- ; (**B**) Rap1- ; and (**C**) Rap2-[ɣ-^{32}P]GTP for 10 minutes at 25 °C. Initial rates (linear for 25 minutes; data not shown) were determined at single time points (10 minutes) and plotted in a bar diagram. Data are mean values (n = 10); error bars are standard errors. Significance was determined by a ordinary one way ANOVA (Uncorrected Fisher's LSD). Mean differs significantly from No Kinase: * $p<0.05$, ** $p<0.01$, *** $p<0.005$, **** $p<0.001$.

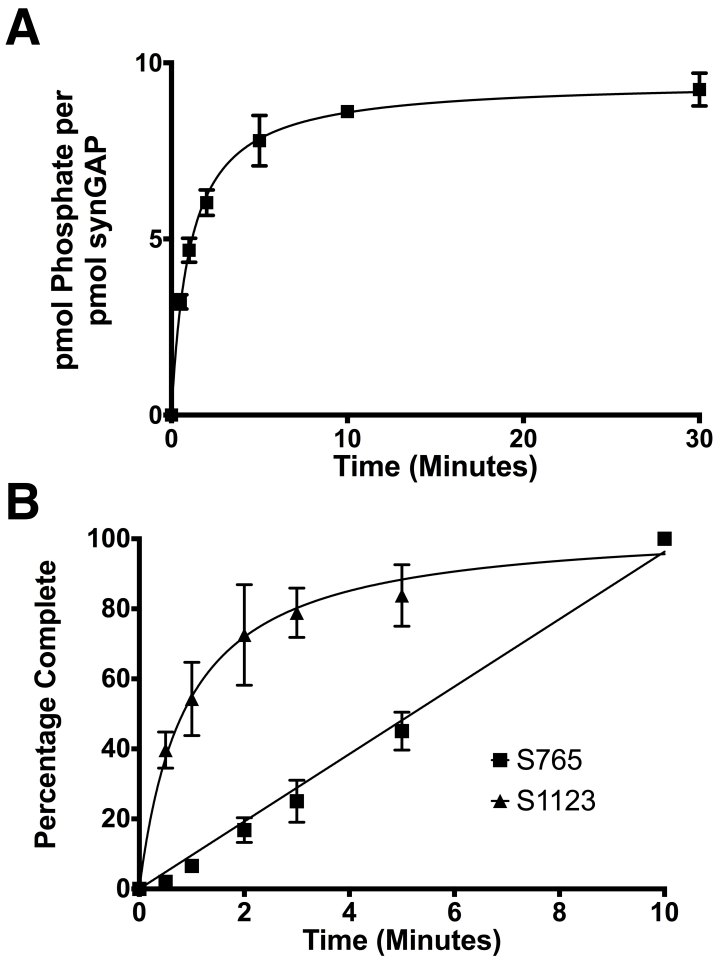


Figure 3.7. Stoichiometry and S765, S1123 Site Occupancy of CaMKII Phosphorylated r-synGAP. (**A**) Stoichiometry of r-synGAP phosphorylation by CaMKII. Soluble, purified r-synGAP (286 nM) was phosphorylated in the presence of purified CaMKII (3.1 nM), 500 μM [γ-^{32}P]-ATP (100 cpm/pmol), 0.7 mM $CaCl_2$, and 3.375 μM CaM, as described under Methods. At the indicated times, reactions were quenched by addition of 4X LDS sample buffer. Radiolabeled r-synGAP was fractionated by SDS-PAGE on 4-12% gels and visualized with a Phosphorimager, as described in Methods. [γ-^{32}P]PO_4^{3-} in the r-synGAP protein band was quantified with the use of ImageQuant TL software, as described in Methods. (**B**) Site occupancy of CaMKII phosphorylated r-synGAP. Soluble, purified r-synGAP (286 nM) was phosphorylated as described in (**A**), using cold ATP in place of [γ-^{32}P]-ATP. Reaction quenching and fractionation by SDS-PAGE was performed as described in (**A**). Following fractionation, r-synGAP was transferred to PVDF membranes and probed with anti-phospho-synGAP-S764/5 and anti-phospho-synGAP-S1123, as described under Methods. Site occupancy was relative to the 10 minute data point. Data are mean values (n = 4); error bars are standard errors.

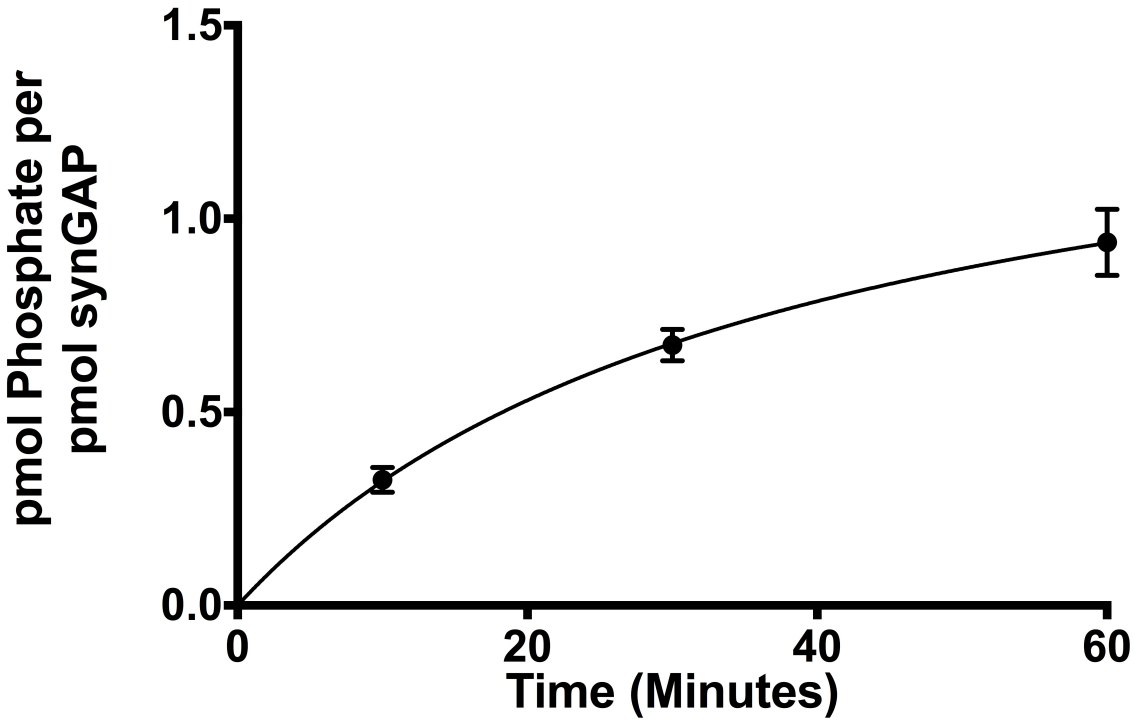


Figure 3.8. Stoichiometry of CDK5 Phosphorylated r-synGAP. Stoichiometry of r-synGAP phosphorylation by CDK5/p35. Soluble, purified r-synGAP (286 nM) was phosphorylated in the presence of purified CDK5/p35 (110 nM) and 500 μM [γ-^{32}P]-ATP (375 cpm/pmol), as described under Methods. At the indicated times, reactions were quenched by addition of 4X LDS sample buffer. Radiolabeled r-synGAP was fractionated by SDS-PAGE on 4-12% gels and visualized with a Phosphorimager as described in Methods. [γ-^{32}P]PO_4^{3-} in the r-synGAP protein band was quantified with the use of ImageQuant TL software, as described under Methods. Data are mean values (n = 4); error bars are standard errors.

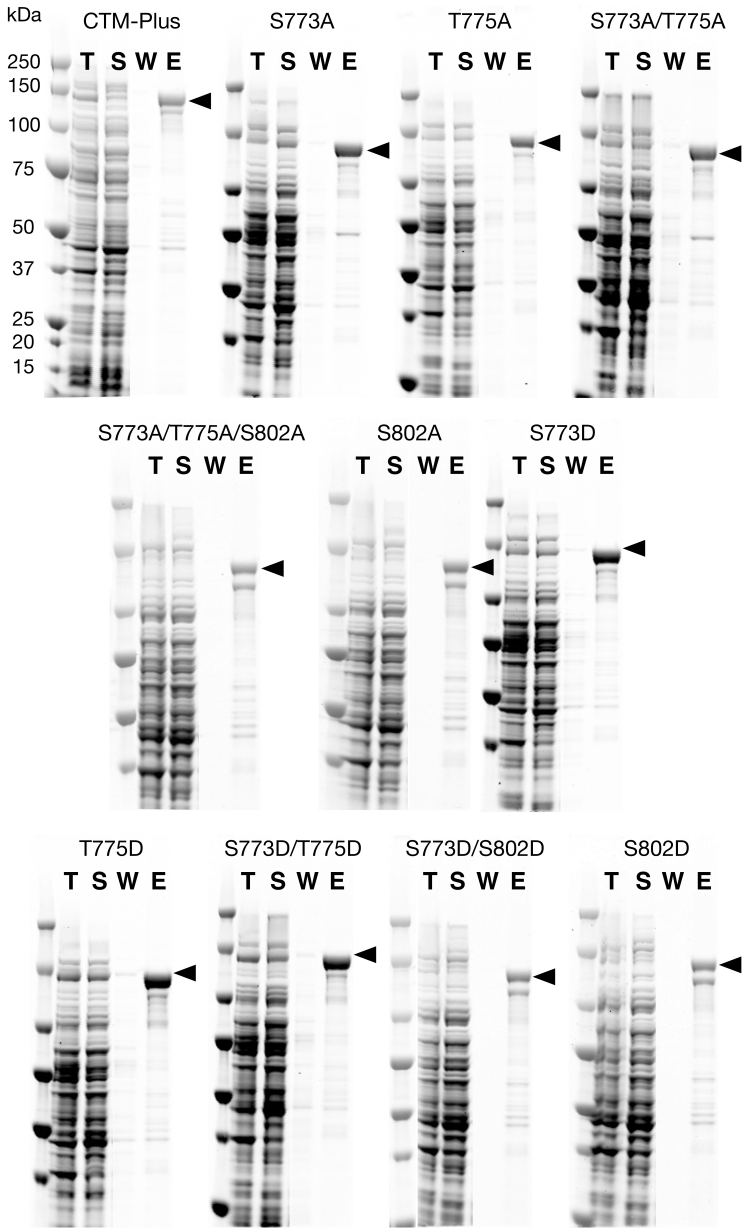


Figure 3.9. Purification of Heterologously Expressed r-synGAP Alanine and Aspartic Acid Mutants. Heterologously expressed 103-1293 r-synGAP alanine and aspartic acid mutants were purified on Talon metal affinity resin, as described in Methods. All proteins are named according to their specific point mutations. The CTM-Plus mutant contains the following point mutations:750A/S751A/S756A/S764A/S765A /S1058A/T1062A/S1064A/S1093A/S1095A/S1099A/S1123A/T1125A. Fractions from each purification and Precision Plus Protein All Blue Standards were separated on a 4-12% gradient SDS-PAGE gel and stained with Gel Code Blue. T, total protein; S, soluble protein; W, wash; E, Talon Co^{2+}-CMA eluate.

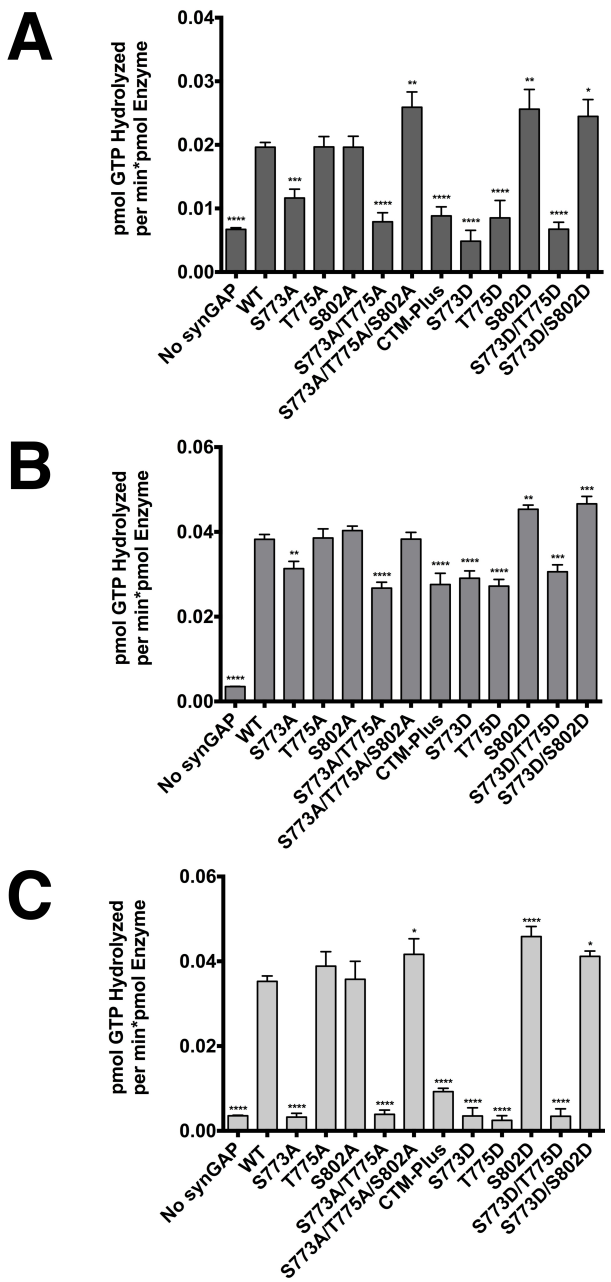


Figure 3.10. Intrinsic GAP Activity of r-synGAP Alanine and Aspartic Acid Mutants. GAP activity of purified r-synGAP (250 nM) mutants assayed with 10 μM (**A**) HRas- ; (**B**) Rap1- ; and (**C**) Rap2-[ɣ-^{32}P]GTP for 10 minutes at 25 °C. Proteins are named according to their specific point mutations, as described in Figure 3.9. Initial rates (linear for 25 minutes; data not shown) were determined at single time points (10 minutes) and plotted in bar diagrams. Data are mean values (n = 5); error bars are standard errors. Significance was determined by a ordinary one way ANOVA (Uncorrected Fisher's LSD). Mean differs significantly from WT: * $p<0.05$, ** $p<0.01$, *** $p<0.005$, **** $p<0.001$.

TABLES

TABLE 3.1. Identification of a CaMKII-Independent Phosphorylation Site in synGAP

Observed m/z	Peptide Mass		Delta	Predicted Phosphopeptide	Residues in synGAP	Identified Site
	Experimental	Calculated				
361.6906	721.3667	721.3411	0.0255	R.LPSPTK.E	771-776	S773

TABLE 3.2. Identification of CaMKII Phosphorylation Sites in r-synGAP

Residues marked with an X are confirmed CaMKII phosphorylation sites in r-synGAP. Modified residues are categorized according to the length of the phosphorylation reaction (2 minutes, 10 minutes) and detection in MS2 or MS3 spectra. Sites highlighted in yellow have phosphorylation false localization rates less than 1% (Performed as described in Methods).

Residue Modified	2 Minutes		10 Minutes	
	MS2	MS3	MS2	MS3
S728			X	X
S737	X	X	X	X
S750			X	
S751	X		X	
S756			X	
S764	X			
S765	X	X	X	X
S808	X		X	
S809	X		X	
S810	X		X	
S821	X		X	
S825	X	X	X	X
S827	X	X	X	X
S842	X			
S843	X			
S882	X	X	X	X
S883	X	X	X	
T885			X	
S892			X	
T897	X	X	X	X
T898	X	X		X
S990	X	X	X	
S1093	X		X	
S1095	X		X	
S1099	X	X	X	X
S1123	X		X	
S1150	X	X	X	
S1171		X	X	X
S1210	X	X	X	X
S1283	X	X	X	X
T1291			X	

TABLE 3.3. Identification of CDK5 Phosphorylation Sites in r-synGAP

Residues marked with an X are confirmed CDK5 phosphorylation sites in r-synGAP. Modified residues are categorized according to the length of the phosphorylation reaction (2 minutes, 10 minutes) and detection in MS2 or MS3 spectra. Sites highlighted in yellow have phosphorylation false localization rates less than 1% (Performed as described in Methods).

Residue Modified	2 Minutes		10 Minutes	
	MS2	MS3	MS2	MS3
S728	X	X	X	X
S773	X	X	X	X
T775	X	X	X	X
S802	X	X	X	X
S842	X		X	X
S1093		X		

Chapter IV

SUBSTRATE DIRECTED REGULATION OF SYNGAP PHOSPHORYLATION BY CALCIUM/CALMODULIN

AUTHORS

Ward G. Walkup IV, Mike J. Sweredoski, Robert L. Graham, Sonja Hess, Mary B. Kennedy*

Division of Biology and Biological Engineering, California Institute of Technology, 1200 East California Blvd, Mail Code 216-76, Pasadena CA 91125, USA

*To whom correspondence should be addressed. Tel: +1 6263953923; Fax: +1 6263958474; Email: kennedym@its.caltech.edu

Key Words:

synGAP, CDK5, PLK2, Calmodulin, Phosphorylation, GTPase, Synaptic Plasticity, Mass Spectrometry

Author Contributions:

WGWIV, MJS & RLG conceived, designed, and carried out experiments; analyzed and interpreted data; and wrote the paper. SH & MBK conceived experiments, interpreted data, and wrote the paper.

AUTHOR'S COMMENTS

This Chapter is a draft of a manuscript destined for submission to the Journal of Biological Chemistry in early 2014. While data collection is nearly complete, several small experiments required for completion of this manuscript are absent from this chapter (see Additional Experiments to Complete Before Manuscript Submission). Once this additional data has been collected, the Results and Discussion section will be significantly expanded in scope.

ABSTRACT

The precise pattern of Ca^{2+} influx through N-methyl-D-aspartate-type glutamate receptors (NMDAR) underlies synaptic plasticity, the ability of excitatory synapses to strengthen or weaken their connections in response to patterns of activity between their connected neurons. Ca^{2+} influx through activated NMDARs results in increased levels of cytosolic Ca^{2+} that can bind to several effector molecules within the spine, the most prominent being calmodulin (CaM), a ubiquitous Ca^{2+} binding protein. Here we show that Ca^{2+} bound CaM (Ca^{2+}/CaM) can bind to a soluble form of recombinant synGAP (r-synGAP), a neuron-specific Ras and Rap GTPase-activating protein (GAP). While Ca^{2+}/CaM binding does not directly affect r-synGAP activity, it does increase the rate and stoichiometry of r-synGAP phosphorylation by the protein kinases cyclin-dependent kinase 5 (CDK5) and polo-like kinase 2 (PLK2).

Phosphorylation of the CDK5 and PLK2 substrates histone H1 and α-casein, respectively, in the presence of Ca^{2+}/CaM does not increase the stoichometry or rate of phosphorylation. The increased rate and stoichiometry of r-synGAP phosphorylation in the presence of Ca^{2+}/CaM is not an intrinsic property of the kinases, and is instead a "substrate directed effect". CDK5 phosphorylation of r-synGAP is increased 2-fold in the presence of Ca^{2+}/CaM, and results in phosphorylation at additional sites not modified in the absence of Ca^{2+}/CaM (S751, S765, T897, S1099, and S1123). PLK2 phosphorylates r-synGAP at several sites, including S750, S751, S756, S765, S825, and T897, that result in an increase of r-synGAPs HRas and Rap1 GAP activity by 70% and 13%, respectively. PLK2 phosphorylation of r-synGAP is increased 2-fold in the presence of Ca^{2+}/CaM, and results in phosphorylation at additional sites not modified in the absence of Ca^{2+}/CaM (S1099, S1123, and S1283).

INTRODUCTION

Memory storage in the brain involves adjustment of the strength of existing synapses and formation of new neural networks. A key process underlying memory formation is synaptic plasticity, the ability of excitatory synapses to strengthen or weaken their connections in response to patterns of activity between their connected neurons [1]. Stronger synapses are defined relatively as synapses that contribute an increased depolarization to a neuronal membrane upon activation, and are thus more

likely to generate an action potential in their postsynaptic neuron. Long-lasting enhancement of synaptic strength elicited by specific patterns of high frequency (Tetanic) synaptic stimulation, resulting in addition of new glutamate receptors and enlargement of the synapse, is referred to as long term potentiation (LTP) [2, 3]. Long-lasting suppression of synaptic strength elicited by specific patterns of long term, low frequency synaptic stimulation, resulting in removal of glutamate receptors and shrinkage of the synapse, is referred to as long term depression (LTD) [4-7]. Although LTP and LTD have opposing effects on the strength of the synapse, they are both controlled by the precise pattern of Ca^{2+} influx through N-methyl-D-aspartate-type glutamate receptors (NMDAR) [2, 8]. High frequency, larger amplitude Ca^{2+} influx through the NMDAR into the postsynaptic neuron promotes the formation of LTP, whereas lower frequency, lower amplitude, and prolonged Ca^{2+} influx through the NMDAR promotes LTD [9, 10].

Ca^{2+} influx through activated NMDARs results in increased levels of cytosolic Ca^{2+} that can bind to several effector molecules within the spine, the most prominent being calmodulin (CaM), a ubiquitous Ca^{2+} binding protein [11-13]. Ca^{2+} bound CaM (Ca^{2+}/CaM) can bind to and activate a multitude of postsynaptic signaling molecules, including Ca^{2+}/calmodulin-dependent protein kinase II (CaMKII) and the phosphatase calcineurin (protein phosphatase 2B). Competition for available Ca^{2+}/CaM can result in the differential activation of CaMKII and calcineurin, which has been hypothesized to determine whether induction of LTP or LTD will occur. Additionally, influx of Ca^{2+} through the NMDAR leads to the the activation of the small GTPases Ras and Rap.

Differential activation of Ras and Rap acts to modulate synaptic strength by promoting the insertion (exocytosis) or removal (endocytosis), respectively, of AMPARs from the synapse [14].

Previous work in Chapter 3 shows that r-synGAP can be phosphorylated by cyclin-dependent kinase 5 (CDK5), a proline directed serine/threonine kinase that has been has been localized to synapses and implicated in the regulation of synaptic plasticity [15, 16]. Here we show that the rate and stoichiometry of r-synGAP phosphorylation by the protein kinases cyclin-dependent kinase 5 (CDK5) and polo-like kinase 2 (PLK2) is increased in the presence of Ca^{2+} bound CaM (Ca^{2+}/CaM). Ca^{2+}/CaM can directly bind to r-synGAP, but does not directly affect its activity. We show that in the presence of Ca^{2+}/CaM phosphorylation of the CDK5 and PLK2 substrates histone H1 and α-casein, respectively, does not increase the stoichometry or rate of phosphorylation. Thus, the increased rate and stoichiometry of r-synGAP phosphorylation in the presence of Ca^{2+}/CaM is not an intrinsic property of the kinases, but is instead a "substrate directed effect". We find that CDK5 phosphorylation of r-synGAP is increased 2-fold in the presence of Ca^{2+}/CaM, and results in phosphorylation at additional sites not modified in the absence of Ca^{2+}/CaM (S751, S765, T897, S1099 and S1123). We also find that PLK2 phosphorylates r-synGAP at several sites including S750, S751, S756, S765, S825, and T897, which result in an increase of r-synGAP's HRas and Rap1 GAP activity by 70% and 13%, respectively. PLK2 phosphorylation of r-synGAP is increased 2-fold in the presence of Ca^{2+}/CaM, and results in phosphorylation at additional sites not modified in the absence of Ca^{2+}/CaM (S1099, S1123, and S1283). As seen with

phosphorylation by CaMKII and CDK5, phosphorylation of r-synGAP by PLK2 may lead to alteration of the ratio of activated Ras and Rap in synapses.

MATERIALS AND METHODS

Unless explicitly mentioned below, all experimental procedures utilized in this chapter were performed exactly as described in Chapter 3.

Stoichiometry and Rate of r-synGAP Phosphorylation by PLK2 or CDK5. Phosphorylation of synGAP by PLK2 (#PV4204, Life Technologies), CDK5/p35 (#14-477M, EMD Millipore), or CDK5/p25 (#14-516, EMD Millipore) was carried out in a reaction mixture containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 0-0.7 mM $CaCl_2$, 0.4 mM EGTA, 500 μM [γ-^{32}P]-ATP (375 cpm/pmol) (6000 Ci/mmol, BLU002Z001MC, Perkin Elmer), 0-3.375 μM calmodulin, 10 mM DTT, 286 nM r-synGAP, and 65.9 PLK2, 110 nM CDK5/p35 or CDK5/p25. The remainder of the assay was performed as described in Chapter 3.

Stoichiometry and Rate of Histone H1 Phosphorylation by CDK5. Phosphorylation of histone H1 (#M2501S, NEB) by CDK5 was carried out in a reaction mixture containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 0-0.7 mM $CaCl_2$, 0.4 mM EGTA, 500 μM [γ-^{32}P]-ATP (375 cpm/pmol), 0-3.375 μM calmodulin, 10 mM DTT, 4.3 μM histone H1 and 110 nM CDK5/p35 or CDK5/p25. The remainder of the assay was performed as described in Chapter 3.

Stoichiometry and Rate of α-Casein Phosphorylation by PLK2. Phosphorylation of dephosphorylated α-casein from bovine milk (#C8032-100MB, Sigma-Aldrich) by PLK2 was carried out in a reaction mixture containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 0-0.7 mM $CaCl_2$, 0.4 mM EGTA, 500 μM [γ-^{32}P]-ATP (375 cpm/pmol), 0-3.375 μM calmodulin, 10 mM DTT, 3.7 μM α-casein, and 65.9 nM PLK2. The remainder of the assay was performed as described in Chapter 3.

Phosphorylation of r-synGAP by PLK2 or CDK5 for use in GAP Assays. Phosphorylation of synGAP by PLK2 or CDK5/p35 for use in GAP assays was carried out in reaction mixtures containing 50 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 0-0.7 mM $CaCl_2$, 0.4 mM EGTA, 30 μM ATP, 0-3.375 μM calmodulin, 10 mM DTT, 724.5 nM r-synGAP, and 230 nM PLK2 or CDK5/p35. The remainder of the assay was performed as described in Chapter 3.

Pull-Down of r-synGAP Using CaM-Sepharose. Rosetta2(DE3) cells expressing 103-725 synGAP or 103-1293 synGAP in pET-47b(+) or empty pET-47b(+) were resuspended in Lysis Buffer, or Lysis Buffer supplemented with 0-5 mM $CaCl_2$ or 0-10 mM EGTA. The resuspended cells were sonicated (Branson, Danbury CT) 2x for 90 seconds/pass (15% power, 1.0 second on, 1.5 seconds off) before clarification by centrifugation at 16,000 × g for 40 min at 4°C. Then 1.7 ml of each of the three sets of clarified lysates (~6 mg/ml) was incubated in batch for 90 minutes at 4°C, with 0.3 ml of

CaM-Sepharose 4B (#17-0529-01, GE Healthcare) or Sepharose 4B (#17-0120-05, GE Healthcare) previously equilibrated in Lysis Buffer or Lysis Buffer supplemented with 5 mM $CaCl_2$ or 10 mM EGTA. Following washing with 40 bed volumes of the respective Lysis Buffer, samples were eluted in Lysis Buffer containing 100 mM EGTA. R-synGAP present in the column eluate was quantified by SDS-PAGE and Western Blotting, as described in Chapter 3.

RESULTS AND DISCUSSION

Effect of Ca^{2+}/CaM on the Stoichiometry and Rate of Phosphorylation of r-synGAP by CDK5. In Chapter 3 we discovered a novel CaMKII-independent phosphorylation site on synGAP (S773) localized to the PSD fraction. The consensus sequence surrounding S773 (SPXR/K) closely resembled the consensus sequence for the proline directed kinase, CDK5. Using mass spectrometry in conjunction with biochemical assays, we showed that soluble, recombinant synGAP (r-synGAP) was a substrate for CDK5, and that phosphorylation of r-synGAP by CDK5 increased its HRas and Rap1 GAP activities by 65% and 20%, respectively. In all biochemical assays performed in Chapter 3, we performed CDK5 phosphorylation reactions in the absence of Ca^{2+}/CaM, because the S773 CDK5 phosphorylation site on synGAP was identified as a CaMKII-independent site. After performing these assays, we considered whether the presence of Ca^{2+}/CaM might affect the rate or stoichiometry of r-synGAP

phosphorylation by CDK5, and if so, if it might affect the GAP activity of r-synGAP.

The presence of Ca^{2+}/CaM affected the overall stoichiometry and rate of phosphorylation of r-synGAP by CDK5/p35 (Fig. 4.1A). When r-synGAP was phosphorylated in the presence of Ca^{2+}/CaM, its overall stoichometry was ~2-fold higher across all compared time points. This effect was limited to reactions containing both Ca^{2+} and CaM. Reactions carried out in the presence of Ca^{2+} or CaM did not exhibit altered phosphorylation rates or stoichiometry relative to reactions carried out in the absence of Ca^{2+} and CaM. These data could have resulted from several different phenomena, including: (i) a substrate driven effect, where Ca^{2+}/CaM binds to and causes a conformational change in r-synGAP that allows for an increase in the rate of its phosphorylation by CDK5/p35, or (ii) an intrinsic property of CDK5/p35, where Ca^{2+}/CaM binding to CDK5/p35 causes a conformational shift that simulates the activity of this kinase toward all substrates.

To determine whether the effects of Ca^{2+}/CaM on the CDK5/p35 catalyzed phosphorylation of r-synGAP are due to an intrinsic property of the CDK5 activator, p35, we phosphorylated r-synGAP with CDK5 bound to p25. The proteolytically activated form of p35, known as p25, lacks the Ca^{2+}/CaM binding site found in p35 and thus is unable to bind Ca^{2+}/CaM [17]. When r-synGAP was phosphorylated by CDK5/p25 (Fig. 4.1B), the rate and overall stoichiometry of phosphorylation was increased in presence of Ca^{2+}/CaM, which is consistent with a substrate driven effect of r-synGAP. The absence of a Ca^{2+}/CaM binding domain in p25 suggests that Ca^{2+}/CaM is not binding to CDK5 or its activators, p35 or p25, and may instead bind directly to r-

synGAP.

Effect of *Ca^{2+}/CaM on the Stoichiometry and Rate of Phosphorylation of Histone H1 by CDK5*. If the increase in r-synGAP phosphorylation by CDK5 is a substrate directed property of r-synGAP, the phosphorylation rates and stoichiometry for other CDK5 substrates should not be affected by the presence of Ca^{2+}/CaM. When histone H1, a substrate of CDK5, was phosphorylated by CDK5/p35 (Fig. 4.2A) or CDK5/p25 (Fig. 4.2B), the rate and overall stoichiometry of phosphorylation was independent of the presence of Ca^{2+}/CaM, suggesting that the effect of Ca^{2+}/CaM is most likely a substrate driven effect of r-synGAP. If p35 binding were responsible for the increased phosphorylation of r-synGAP, we should have observed an increase in the rate and stoichiometry of histone H1 phosphorylation in the presence of Ca^{2+}/CaM for p35 relative to p25.

Effect of Ca^{2+}/CaM on the Stoichiometry and Rate of Phosphorylation of r-synGAP by PLK2. Membrane bound synGAP has previously been shown to be a substrate for the kinase polo-like kinase 2 (PLK2) [18]. In order to reduce the chances of the enhanced phosphorylation rate being due to an as of yet undiscovered property of CDK5/p35/p25, we chose to phosphorylate r-synGAP with PLK2 in the presence or absence of Ca^{2+}/CaM. When r-synGAP is phosphorylated by PLK2 in the presence of Ca^{2+}/CaM, its overall stoichometry and rate of phosphorylation is doubled across all compared time points (Fig. 4.3). This effect was limited to reactions containing both

Ca^{2+} and CaM. Reactions carried out in the presence of Ca^{2+} or CaM alone did not exhibit altered phosphorylation rates or stoichiometry relative to reactions carried out in the absence of Ca^{2}/CaM. As with phosphoryation of r-synGAP by CDK5, the increased rate of r-synGAP phosphorylation by PLK2 in the presence of Ca^{2+}/CaM seems to be a substrate directed effect of r-synGAP.

Effect of Ca^{2+}/CaM on the Stoichiometry and Rate of Phosphorylation of α-Casein by PLK2. If the increase in r-synGAP phosphorylation by PLK2 is a substrate directed property of r-synGAP, the phosphorylation rates and stoichiometry of other PLK2 substrates should not be affected by the presence of Ca^{2+}/CaM. Dephosphorylated α-casein is a non-physiological substrate of PLK2 that has been used to determine the kinetics of PLK2 activation. If Ca^{2+}/CaM binding to r-synGAP is a substrate driven effect unique to r-synGAP, one would anticipate that the rate and stoichiometry of α-casein phosphorylation would be independent of Ca^{2+}/CaM. When α-casein was phosphorylated by PLK2 (Fig. 4.4) , the rate and overall stoichiometry of phosphorylation was unchanged in the presence or absence of Ca^{2+}/CaM. We conclude that increased phosphorylation of r-synGAP in the presence of Ca^{2+}/CaM is a substrate driven effect of r-synGAP, and may be mediated by direct binding of Ca^{2+}/CaM to r-synGAP.

r-synGAP Binds to CaM-Sepharose in the Presence of Ca^{2+}. We hypothesized that the increased rate and stoichiometry of r-synGAP phosphorylation in the presence of

Ca^{2+}/CaM may be due to a conformational shift in r-synGAP structure caused by binding to Ca^{2+}/CaM. To determine if r-synGAP binds Ca^{2+}/CaM, we tested the binding of 103-1293 r-synGAP to CaM immobilized on sepharose-4B beads (Fig. 4.5) in the presence and absence of Ca^{2+}.

r-synGAP bound to CaM immobilized on Sepharose 4B in the presence of Ca^{2+}. When EGTA was added to the buffer, almost all binding was prevented. r-synGAP did not bind to Sepharose 4B that lacks NHS coupled CaM. 103-725 synGAP bound more weakly to Ca^{2+}/CaM Sepharose 4B than 103-1293 synGAP (data not shown), presumably because of either additional Ca^{2+}/CaM binding sites or additional structural features in the disordered, coiled-coil, or PDZ binding domains.

Effect of Ca^{2+}/CaM on CDK5 Phosphorylation Sites in r-synGAP. In Chapter 3 we showed that CDK5/p35 can phosphorylate r-synGAP at two primary sites (S773/T775 and S802) and two additional sites (S728, S842) in the absence of Ca^{2+}/CaM. Using mass spectrometry (as described in Methods from Chapter 3), we found that CDK5 phosphorylation of r-synGAP in the presence of Ca^{2+}/CaM resulted in phosphorylation at several sites previously classified as CaMKII sites (S751, S765, T897, S1093, S1099, and S1123) (Table 4.1). CDK5 phosphorylation reactions with r-synGAP in the presence of Ca^{2+}/CaM and mass spectrometry were carried out as described in the Methods section of Chapter 3. None of the Ca^{2+}/CaM dependent CDK5 phosphorylation sites identified in r-synGAP (S751, S765, S843, S844, T897, S1093, S1099, and S1123) are located in CDK5 consensus sequences (S/TP motifs).

While S751, S765, T897, S1093, S1099, and S1123 exhibited a phosphorylation false localization rate of less than 1%, S843 and S844 had rates greater than 5%. Due to their proximity to S842 and low scoring mass spectra, we considered the latter two residues to be false positive identifications. Interestingly, it appears that S751, S765, T897, S1093, S1099, and S1123 may not be CaMKII specific sites (S1093 and S1123 are located in R/KXXS*/T* CaMKII consensus sequences), and instead may be Ca^{2+}/CaM dependent sites of phosphorylation.

Effect of Ca^{2+}/CaM on PLK2 Phosphorylation Sites in r-synGAP. synGAP has previously been shown to be phosphorylated at residues S379, S385, S449, S466, S836, S840, S842, and S895 by PLK2 [18]. Of these previously identified PLK2 phosphorylation sites in synGAP, only S385, S449, S840, and S842 have been shown to have an effect on GAP activity [18]. When r-synGAP was phosphorylated by PLK2 in the absence of Ca^{2+}/CaM, it was found to be phosphorylated at residues S750, S751, S756, S765, S808, S810, S821, S825, S827, and T897 (Table 4.2). Phosphorylation at sites S750, S751, S756, and S765 occurred at both the 2 minute and 10 minute time points, whereas phosphorylation at S808, S810, S821, S825, S827, and T897 occurred only at 10 minutes. S750, S751, S756 and S765, S825, and T897 all exhibited phosphorylation false localization rates of less than 1% and were detected in the PLK2 phosphorylation reactions conducted in the presence and absence of Ca^{2+}/CaM. None of PLK2 phosphorylation sites in r-synGAP identified in this study were previously identified by Lee *et al.* [18]. It is likely that these discrepancies arise from differences

in the methods of preparation of synGAP and PLK2: purified, soluble r-synGAP constructs and PLK2 were utilized in this study, while [18] used cell lysates from COS cells overexpressing synGAP and PLK2.

In the presence of Ca^{2+}/CaM, PLK2 phosphorylates r-synGAP at several additional sites, including S737, S1099, S1123, and S1283 (Table 4.2). All four sites exhibited high scoring Mascot scores and phosphorylation false localization rates of less than 1%. Like CDK5 phosphorylation reactions carried out in the presence of Ca^{2+}/CaM, the presence of Ca^{2+}/CaM in PLK2 phosphorylation reactions resulted in r-synGAP phosphorylation at residues previously classified as CaMKII sites (S737, S1099, S1123, and S1283). Based on this data, we believe that Ca^{2+}/CaM binds to and causes a conformational change in r-synGAP that exposes previously inaccessible residues for phosphorylation.

Activity of r-synGAP in the Presence of Ca^{2+}/CaM. As stated above, Ca^{2+}/CaM can bind to and activate a multitude of postsynaptic signaling molecules, including CaMKII and the protein phosphatase, calcineurin. Because binding to Ca^{2+}/CaM is a common regulatory mechanism for protein activation in synapses, we tested whether the activity of r-synGAP is regulated by binding to Ca^{2+}/CaM. To determine if r-synGAP activity is regulated by binding to Ca^{2+}/CaM, we measured the GAP activity of r-synGAP in the presence and absence of Ca^{2+}/CaM, as described in Methods (Fig. 4.6). Binding of Ca^{2+}/CaM to r-synGAP did not directly affect its Ras or Rap GAP activity. We next sought to determine if binding of Ca^{2+}/CaM to r-synGAP can indirectly affect r-

synGAP's activity.

Activity of r-synGAP After Phosphorylation by CDK5 in the Presence of Ca^{2+}/CaM. To determine if CDK5 phosphorylation in the presence of Ca^{2+}/CaM affects r-synGAP's activity, we measured the GAP activity of r-synGAP before and after CDK5/p35 phosphorylation in the presence and absence of Ca^{2+}/CaM, as described in Methods (Fig. 4.7). As shown in Chapter 3, phosphorylation by CDK5/p35 in the absence of Ca^{2+}/CaM increased r-synGAP's HRas and Rap1 GAP activity but not its Rap2 GAP activity. Phosphorylation of synGAP by CDK5/p35 in the presence of Ca^{2+}/CaM had no effect on Rap (Rap1 and Rap2) GAP activity, relative to unphosphorylated r-synGAP, which was unexpected given the 20% increase in Rap1 GAP activity in the absence of Ca^{2+}/CaM (HRas activity in presence of Ca^{2+}/CaM was not measured due to experimental error). Phosphorylation of r-synGAP by CDK5/p35 in the presence of Ca^{2+}/CaM may have resulted in increased phosphorylation at S773 relative to S802, or at the additional phosphorylated sites (S751, S765, T897, S1099, and/or S1123), and may alter r-synGAP's activity in a non-trivial manner.

Activity of r-synGAP After Phosphorylation by PLK2 in the Presence of Ca^{2+}/CaM. It has been previously reported that the Ras GAP activity of synGAP is activated by phosphorylation with PLK2 [18]. To determine if PLK2 phosphorylation in the presence or absence of Ca^{2+}/CaM affects r-synGAP's activity, we measured the GAP activity of r-synGAP before and after PLK2 phosphorylation, as described in Methods

(Fig. 4.8). Phosphorylation of r-synGAP by PLK2 in the presence or absence Ca^{2+}/CaM increased r-synGAP's HRas and Rap1 GAP activity. r-synGAP's RasGAP activity was increased by 69-72% and Rap1GAP by 11-14%. PLK2 phosphorylation of r-synGAP produced a steady increase in HRas and Rap1 GAP activity with phosphorylation time, mirroring the reaction profile of PLK2 phosphorylation.

Phosphorylation of r-synGAP by PLK2 in the absence of Ca^{2+}/CaM did not affect Rap2 GAP activity; however, phosphorylation by PLK2 in the presence of Ca^{2+}/CaM produced an 11% increase in r-synGAP's Rap2 GAP activity. The increase in Rap2 GAP activity in the presence of Ca^{2+}/CaM may be due to phosphorylation at S1099, S1123, and S1283 residues.

It has previously been shown that PLK2 and CDK5 work in concert to target the postsynaptic RapGAP SPAR for degradation [19, 20]. Given that the effects of PLK2 and CDK5 phosphorylation on r-synGAP activity are very similar (both increase Ras GAP activity by ~70% and Rap1 GAP by 10-20%), PLK2 and CDK5 may act in concert to differentially regulate r-synGAP activity.

ADDITIONAL EXPERIMENTS TO COMPLETE BEFORE MANUSCRIPT SUBMISSION

RasGAP Activity of r-synGAP After Phosphorylation by CDK5 in the Presence of Ca^{2+}/CaM. An experimental error during collection of this data precluded its

incorporation into this chapter. This experiment will be repeated as described in Chapter 3.

Determine K_D of r-synGAP - Ca^{2+}/CaM Interaction. The results of initial Biacore experiments suggested that r-synGAP bound Ca^{2+}/CaM with a K_D in the nanomolar range (data not shown). Unfortunately, we were unable to reproduce these data due to technical difficulties. Technical difficulties included the binding and subsequent precipitation of Ca^{2+} by Polysorbate 20 (Tween-20) present in 1x HBS-P+ Buffer, malfunctioning of the Biacore T-200 sample compartment temperature control (synGAP and CaM are temperature sensitive) and non-specific adsorption of synGAP and CaM to the Biacore CM5 chip. We believe that we have resolved these technical difficulties and will repeat the Biacore experiments in December, 2013. If the Biacore experiments prove to be intractable, we will measure the affinity of the Ca^{2+}/CaM - r-synGAP interaction by ligand binding assays utilizing bead-bound r-synGAP and a dilution series of Ca^{2+}/CaM in solution, or vice versa.

REFERENCES

[1] T.V. Bliss, T. Lomo, Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. The Journal of physiology 232 (1973) 331-356.

[2] T.V. Bliss, G.L. Collingridge, A synaptic model of memory: long-term potentiation in the hippocampus. Nature 361 (1993) 31-39.

[3] R.C. Malenka, R.A. Nicoll, Long-term potentiation--a decade of progress? Science (New York, NY) 285 (1999) 1870-1874.

[4] S.M. Dudek, M.F. Bear, Homosynaptic long-term depression in area CA1 of hippocampus and effects of N-methyl-D-aspartate receptor blockade. Proceedings of the National Academy of Sciences of the United States of America 89 (1992) 4363-4367.

[5] R.M. Mulkey, S. Endo, S. Shenolikar, R.C. Malenka, Involvement of a calcineurin/inhibitor-1 phosphatase cascade in hippocampal long-term depression. Nature 369 (1994) 486-488.

[6] R.M. Mulkey, C.E. Herron, R.C. Malenka, An essential role for protein phosphatases in hippocampal long-term depression. Science (New York, NY) 261 (1993) 1051-1055.

[7] R.M. Mulkey, R.C. Malenka, Mechanisms underlying induction of homosynaptic long-term depression in area CA1 of the hippocampus. Neuron 9 (1992) 967-975.

[8] M.B. Kennedy, Signal-processing machines at the postsynaptic density. Science (New York, NY) 290 (2000) 750-754.

[9] K.M. Franks, T.J. Sejnowski, Complexity of calcium signaling in synaptic spines. BioEssays : news and reviews in molecular, cellular and developmental biology 24 (2002) 1130-1144.

[10] P.J. Sjostrom, S.B. Nelson, Spike timing, calcium signals and synaptic plasticity. Current opinion in neurobiology 12 (2002) 305-314.

[11] D. Chin, A.R. Means, Calmodulin: a prototypical calcium sensor. Trends in cell biology 10 (2000) 322-328.

[12] S. Linse, A. Helmersson, S. Forsen, Calcium binding to calmodulin and its globular domains. The Journal of biological chemistry 266 (1991) 8050-8054.

[13] C.B. Klee, Interaction of CaM with calcium and target proteins., Elsevier, Amsterdam, 1988.

[14] J.J. Zhu, Y. Qin, M. Zhao, L. Van Aelst, R. Malinow, Ras and Rap control AMPA receptor trafficking during synaptic plasticity. Cell 110 (2002) 443-455.

[15] M. Angelo, F. Plattner, K.P. Giese, Cyclin-dependent kinase 5 in synaptic plasticity, learning and memory. Journal of neurochemistry 99 (2006) 353-370.

[16] Z. Fu, S.H. Lee, A. Simonetta, J. Hansen, M. Sheng, D.T. Pak, Differential roles of Rap1 and Rap2 small GTPases in neurite retraction and synapse elimination in hippocampal spiny neurons. Journal of neurochemistry 100 (2007) 118-131.

[17] L. He, Z. Hou, R.Z. Qi, Calmodulin binding and Cdk5 phosphorylation of p35 regulate its effect on microtubules. The Journal of biological chemistry 283 (2008) 13252-13260.

[18] K.J. Lee, Y. Lee, A. Rozeboom, J.Y. Lee, N. Udagawa, H.S. Hoe, D.T. Pak, Requirement for Plk2 in orchestrated ras and rap signaling, homeostatic structural plasticity, and memory. Neuron 69 (2011) 957-973.

[19] D.P. Seeburg, D. Pak, M. Sheng, Polo-like kinases in the nervous system. Oncogene 24 (2005) 292-298.

[20] D.P. Seeburg, M. Feliu-Mojer, J. Gaiottino, D.T. Pak, M. Sheng, Critical role of CDK5 and Polo-like kinase 2 in homeostatic synaptic plasticity during elevated activity. Neuron 58 (2008) 571-583.

FIGURES

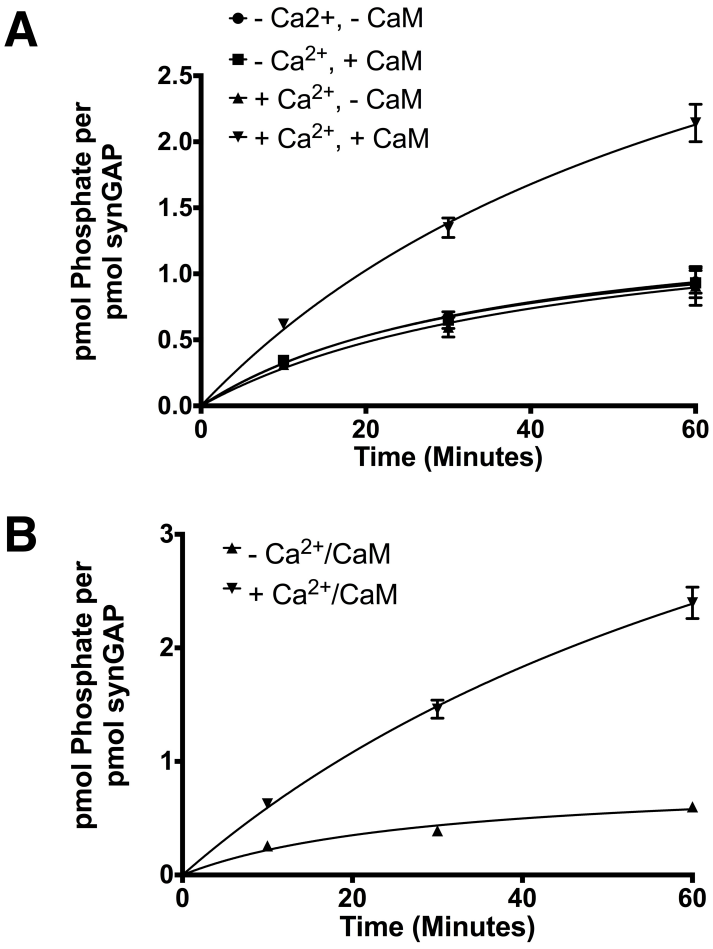


Figure 4.1: Effect of Ca^{2+}/CaM on Stoichiometry of CDK5 Phosphorylation of r-synGAP. Stoichiometry of r-synGAP phosphorylation by (**A**) CDK5/p35 and (**B**) CDK5/p25 in the presence or absence of Ca^{2+}/CaM. Soluble, purified r-synGAP (286 nM) was phosphorylated in the presence of purified CDK5/p35 or CDK5/p25 (110 nM), 500 μM [γ-^{32}P]-ATP (375 cpm/pmol), 0-0.7 mM $CaCl_2$, and 0-3.375 μM CaM as described under Methods. At the indicated times, reactions were quenched by addition of 4X LDS sample buffer. Radiolabeled r-synGAP was fractionated by SDS-PAGE on 4-12% gels and visualized with a Phosphorimager as described in Methods. [γ-^{32}P]PO_4^{3-} in the r-synGAP protein band was quantified with the use of ImageQuant TL software, as described under Methods. Data are mean values (n = 4); error bars are standard errors.

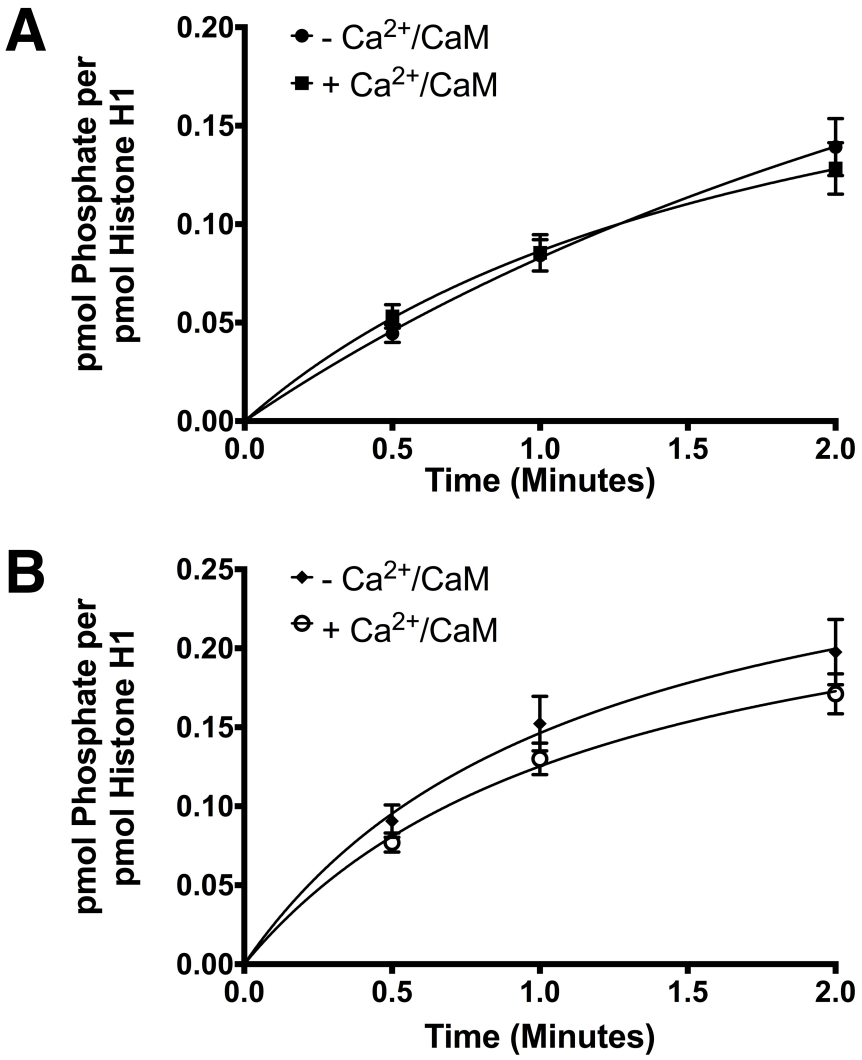


Figure 4.2: Effect of Ca^{2+}/CaM on Stoichiometry of CDK5 Phosphorylation of Histone H1. Stoichiometry of histone H1 phosphorylation by (**A**) CDK5/p35 and (**B**) CDK5/p25 in the presence or absence of Ca^{2+}/CaM. Soluble, purified histone H1 (4.3 μM) was phosphorylated in the presence of purified CDK5/p35 or CDK5/p25 (110 nM), 500 μM [γ-^{32}P]-ATP (375 cpm/pmol), 0-0.7 mM $CaCl_2$, and 0-3.375 μM CaM as described under Methods. At the indicated times, reactions were quenched by addition of 4X LDS sample buffer. Radiolabeled histone H1 was fractionated by SDS-PAGE on 4-12% gels and visualized with a Phosphorimager as described in Methods. [γ-^{32}P]PO_4^{3-} in the histone H1 protein band was quantified with the use of ImageQuant TL software, as described under Methods. Data are mean values (n = 4); error bars are standard errors.

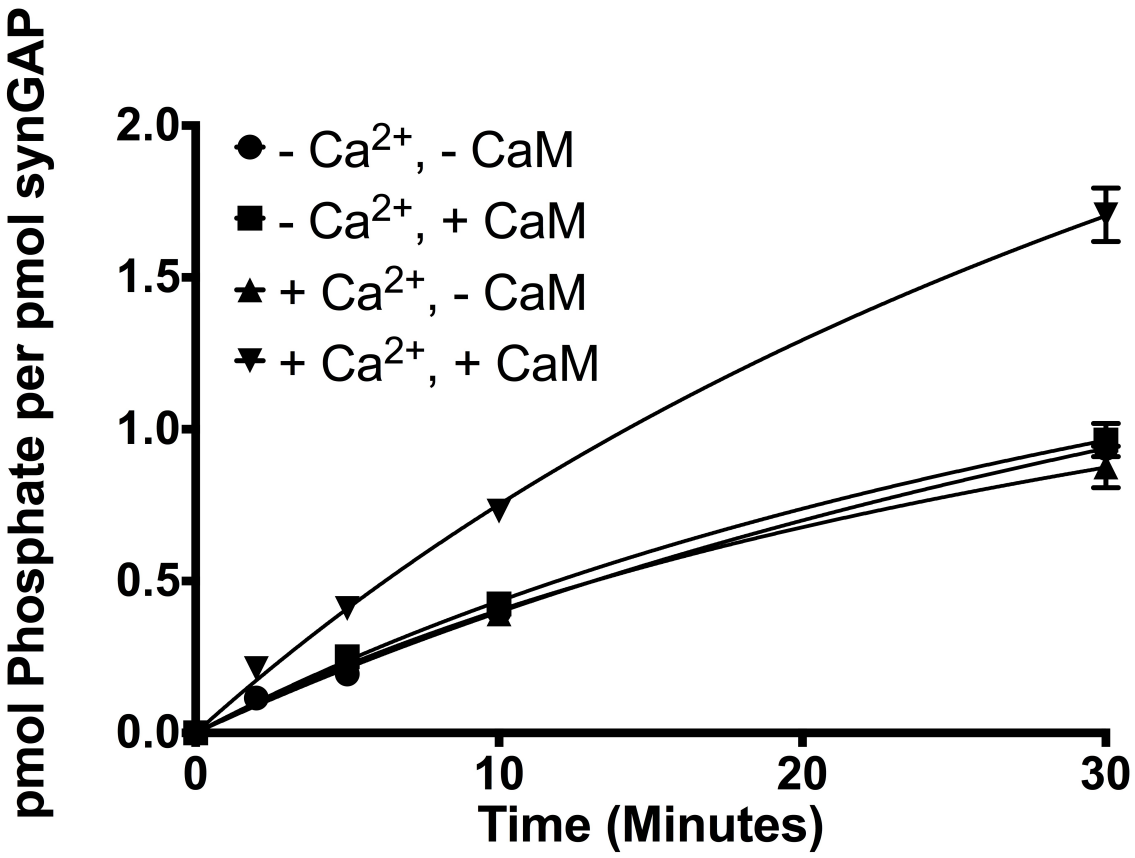


Figure 4.3: Effect of Ca^{2+}/CaM on Stoichiometry of PLK2 Phosphorylation of r-synGAP. Stoichiometry of r-synGAP phosphorylation by PLK2 in the presence or absence of Ca^{2+}/CaM. Soluble, purified r-synGAP (286 nM) was phosphorylated in the presence of purified PLK2 (65.9 nM), 500 μM [γ-^{32}P]-ATP (375 cpm/pmol), 0-0.7 mM $CaCl_2$, and 0-3.375 μM CaM as described under Methods. At the indicated times, reactions were quenched by addition of 4X LDS sample buffer. Radiolabeled r-synGAP was fractionated by SDS-PAGE on 4-12% gels and visualized with a Phosphorimager as described in Methods. [γ-^{32}P]PO_4^{3-} in the r-synGAP protein band was quantified with the use of ImageQuant TL software, as described under Methods. Data are mean values (n = 4); error bars are standard errors.

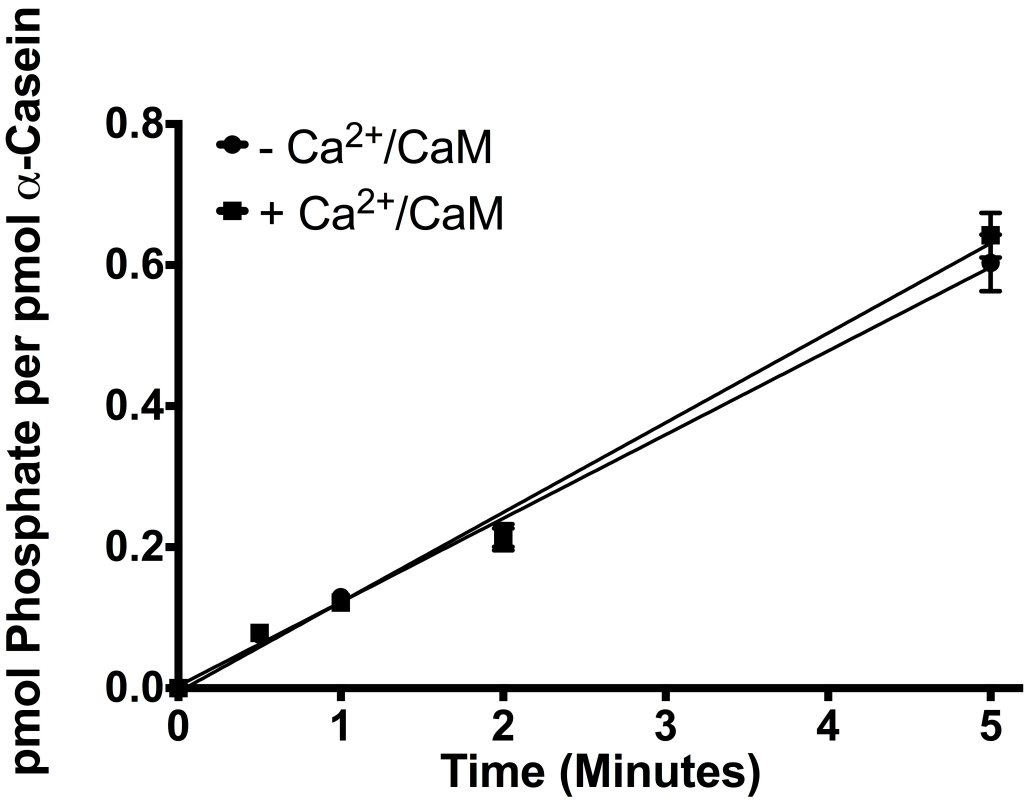


Figure 4.4: Effect of Ca^{2+}/CaM on Stoichiometry of PLK2 Phosphorylation of α-Casein. Stoichiometry of α-casein phosphorylation by PLK2 in the presence or absence of Ca^{2+}/CaM. Soluble, purified α-casein (3.7 μM) was phosphorylated in the presence of purified PLK2 (65.9 nM), 500 μM [γ-^{32}P]-ATP (375 cpm/pmol), 0-0.7 mM $CaCl_2$, and 0-3.375 μM CaM as described under Methods. At the indicated times, reactions were quenched by addition of 4X LDS sample buffer. Radiolabeled α-casein was fractionated by SDS-PAGE on 4-12% gels and visualized with a Phosphorimager as described in Methods. [γ-^{32}P]PO_4^{3-} in the α-casein protein band was quantified with the use of ImageQuant TL software, as described under Methods. Data are mean values (n = 4); error bars are standard errors.

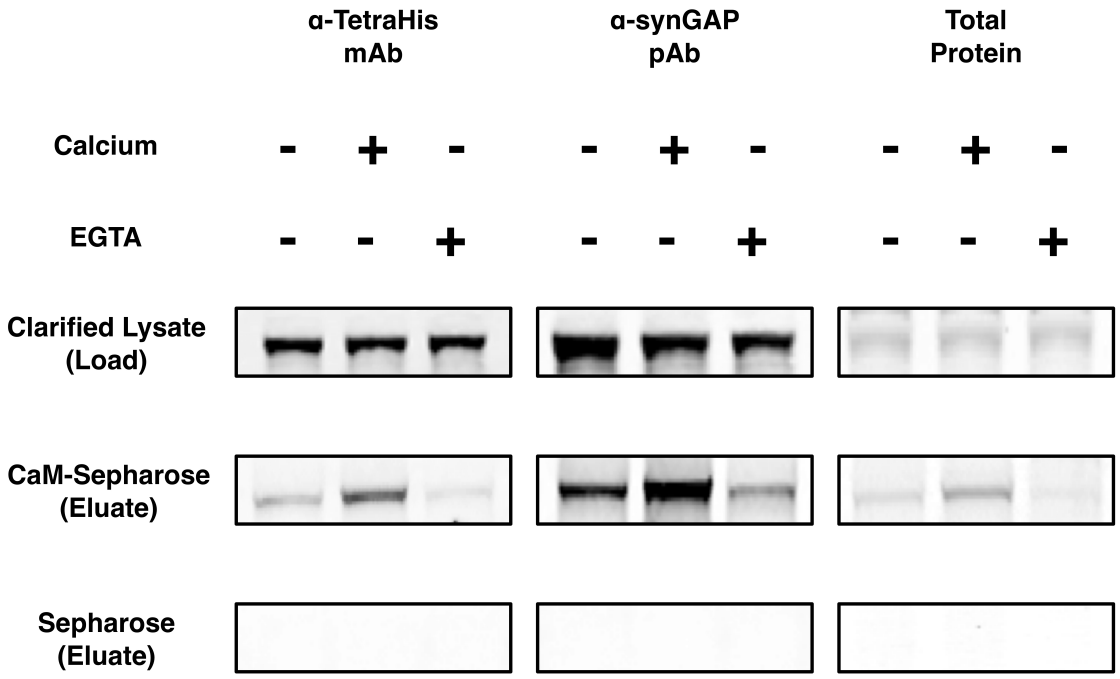


Figure 4.5: r-synGAP Binds to CaM-Sepharose in the Presence of Ca^{2+}. 1.7 ml of *E. coli* lysate containing r-synGAP (103-1293) was incubated with 300 µl of CaM-Sepharose 4B or control Sepharose 4B in the presence or absence of 5 mM $CaCl_2$ or 10 mM EGTA, as described in Methods. Fractions from each pulldown and Precision Plus Protein All Blue Standards were separated on a 4-12% gradient SDS-PAGE gel and stained with Gel Code Blue, or visualized after Western Blotting with anti-TetraHis monoclonal antibody or anti-synGAP polyclonal antibody.

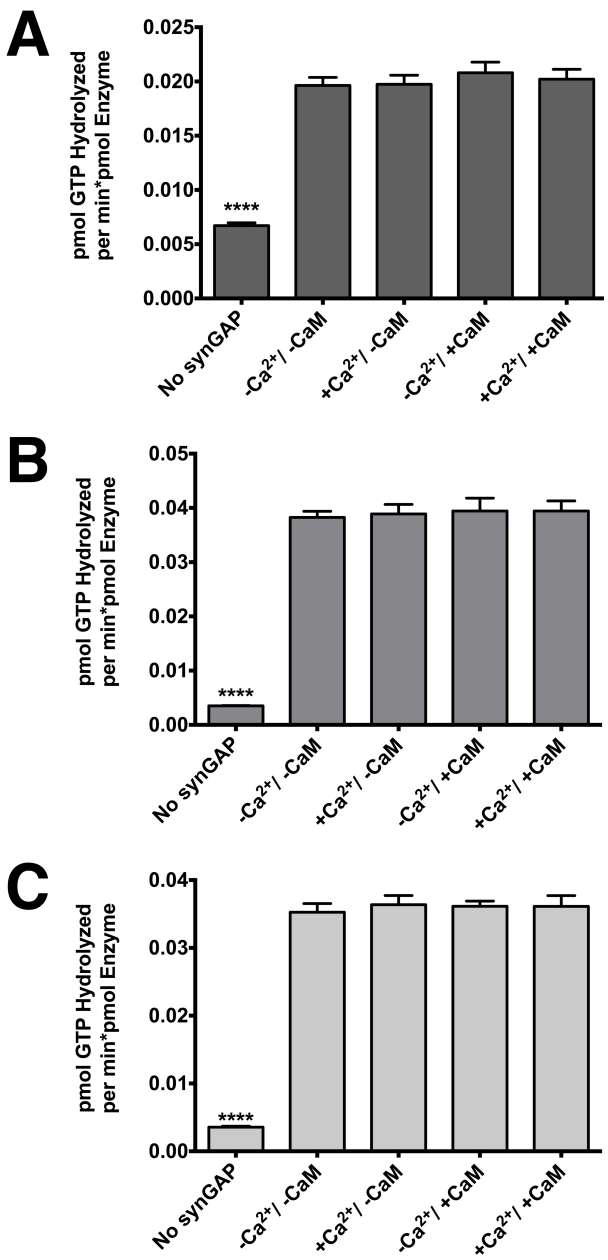


Figure 4.6. Effect of Ca^{2+}/CaM Binding on GAP Activity of r-synGAP. GAP activity of purified r-synGAP (250 nM) in the presence or absence of 3 μM Ca^{2+}/CaM was assayed with 10 μM (**A**) HRas- ; (**B**) Rap1- ; and (**C**) Rap2-[γ-^{32}P]GTP for 10 minutes at 25 °C. Initial rates (Linear for 25 minutes; data not shown) were determined at single time points (10 minutes) and plotted in bar diagrams. Data are mean values (n = 4); error bars are standard errors. Differs from the mean for -Ca^{2+}/-CaM: * $p<0.05$, ** $p<0.01$, *** $p<0.005$, **** $p<0.001$.

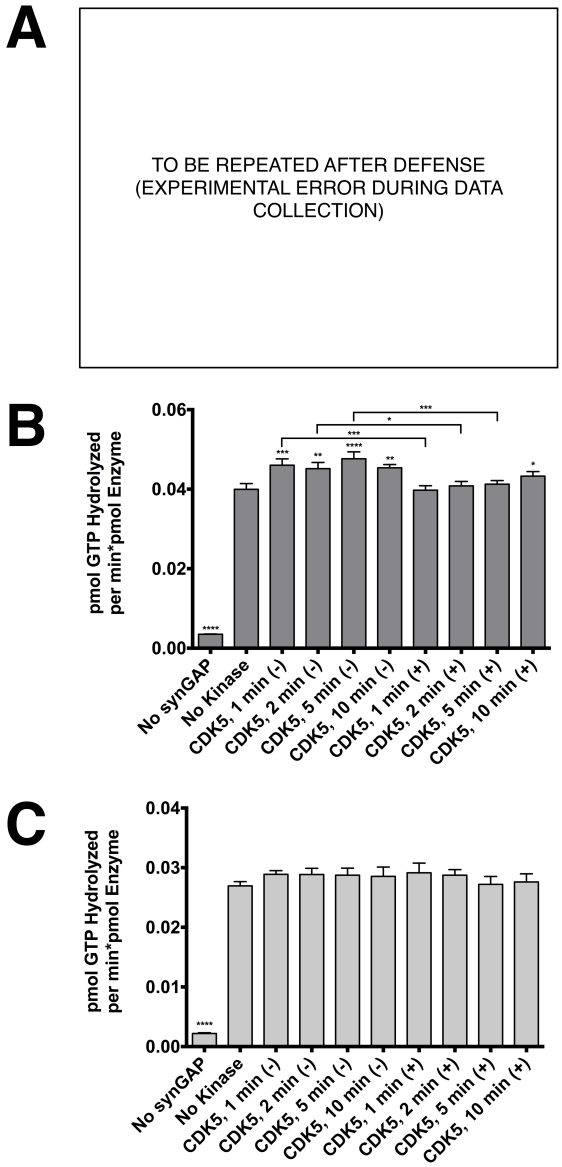


Figure 4.7. GAP Activity of r-synGAP After Phosphorylation by CDK5 in the Presence of Ca^{2+}/CaM. GAP activity of purified r-synGAP (250 nM) before and after phosphorylation by CDK5/p35 for 1 to 10 minutes in the presence or absence of 3 μM Ca^{2+}/CaM was assayed with 10 μM (**A**) HRas- ; (**B**) Rap1- ; and (**C**) Rap2-[γ-^{32}P]GTP for 10 minutes at 25 °C. Initial rates (Linear for 25 minutes; data not shown) were determined at single time points (10 minutes) and plotted in bar diagrams. Data are mean values (n = 10); error bars are standard errors. Differs from the mean for No Kinase: * $p<0.05$, ** $p<0.01$, *** $p<0.005$, **** $p<0.001$. Statistical data shown above the long horizontal brackets in the figure indicate significant differences between means for the data at the two ends of the long bracket.

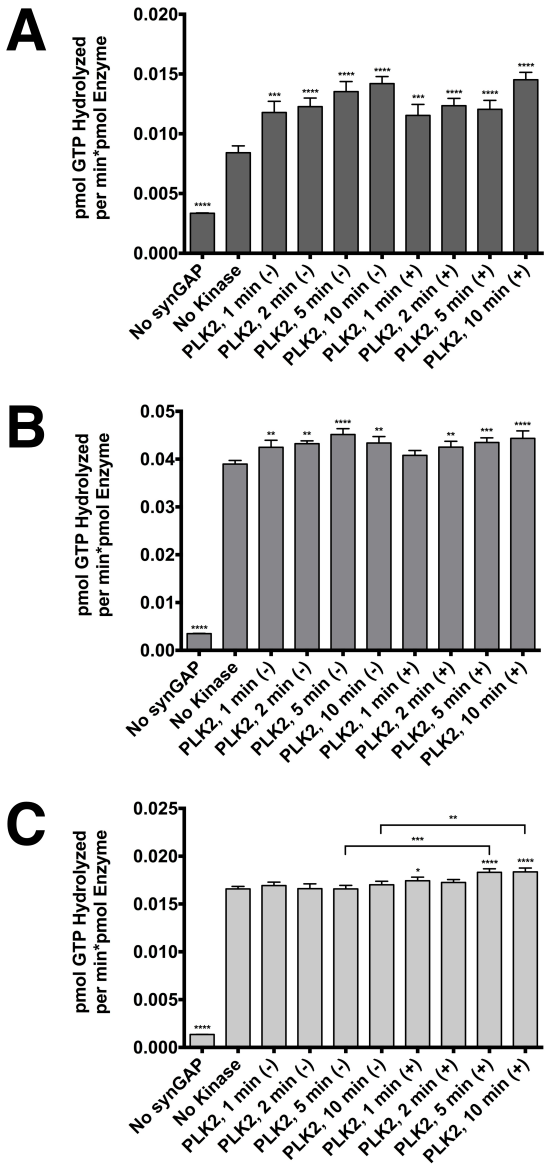


Figure 4.8. GAP Activity of r-synGAP After Phosphorylation by PLK2 in the Presence of Ca^{2+}/CaM. GAP activity of purified r-synGAP (250 nM) before and after phosphorylation by PLK2 for 1 to 10 minutes in the presence (+) or absence (-) of 3 μM Ca^{2+}/CaM was assayed with 10 μM (**A**) HRas- ; (**B**) Rap1- ; and (**C**) Rap2-[γ-^{32}P]GTP for 10 minutes at 25 °C. Initial rates (Linear for 25 minutes; data not shown) were determined at single time points (10 minutes) and plotted in a bar diagram. Data are mean values (n = 10); error bars are standard errors. Differs from the mean for No Kinase: * $p<0.05$, ** $p<0.01$, *** $p<0.005$, **** $p<0.001$. Statistical data shown above the long horizontal brackets in the figure indicate significant differences between means for the data at the two ends of the long bracket.

TABLES

TABLE 4.1. Identification of CDK5 Phosphorylation Sites in r-synGAP

Residues marked with an X are confirmed CDK5 phosphorylation sites in r-synGAP. Modified residues are categorized according to the length of the phosphorylation reaction (2 minutes, 10 minutes), presence of Ca^{2+}/CaM in the phosphorylation reaction, and detection in MS2 or MS3 spectra. Sites highlighted in yellow have phosphorylation false localization rates of less than 1% (performed as described in Methods).

Residue Modified	-Ca^{2+}/CaM				+Ca^{2+}/CaM			
	2 Minutes		10 Minutes		2 Minutes		10 Minutes	
	MS2	MS3	MS2	MS3	MS2	MS3	MS2	MS3
S728	X	X	X	X	X	X	X	X
S751							X	X
S765							X	
S773	X	X	X	X	X	X	X	X
T775	X	X	X	X			X	X
S802	X	X	X	X	X	X	X	X
S810							X	
S842	X		X	X	X		X	X
S843							X	
S844							X	X
T897					X	X	X	X
S1093		X					X	
S1099							X	X
S1123							X	

TABLE 4.2. Identification of PLK2 Phosphorylation Sites in r-synGAP

Residues marked with an X are confirmed PLK2 phosphorylation sites in r-synGAP. Modified residues are categorized according to the length of the phosphorylation reaction (2 minutes, 10 minutes), presence of Ca^{2+}/CaM in the phosphorylation reaction, and detection in MS2 or MS3 spectra. Sites highlighted in yellow have phosphorylation false localization rates of less than 1% (performed as described in Methods).

Residue Modified	-Ca^{2+}/CaM				+Ca^{2+}/CaM			
	2 Minutes		10 Minutes		2 Minutes		10 Minutes	
	MS2	MS3	MS2	MS3	MS2	MS3	MS2	MS3
S737							X	
S750	X	X	X	X	X	X	X	X
S751	X		X	X	X		X	
S756	X							
S765	X		X		X		X	
S808	X		X		X		X	
S810			X				X	
S821			X					
S825			X	X			X	X
S827			X				X	X
T897			X	X	X	X	X	X
S1099							X	
S1123					X		X	
S1283					X	X	X	X

A p p e n d i x I

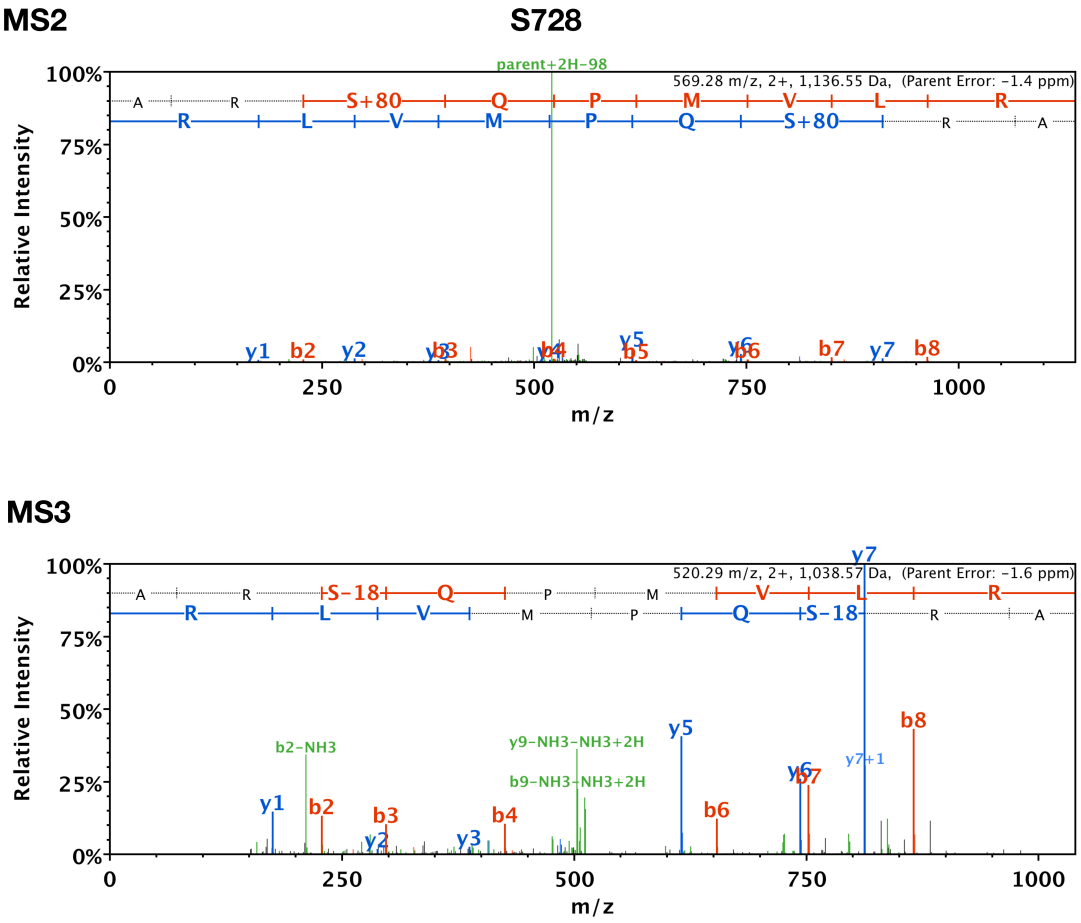
MASS SPECTRA FOR CAMKII, CDK5 AND PLK2 PHOSPHORYLATION SITES IN SYNGAP

AUTHOR'S COMMENTS

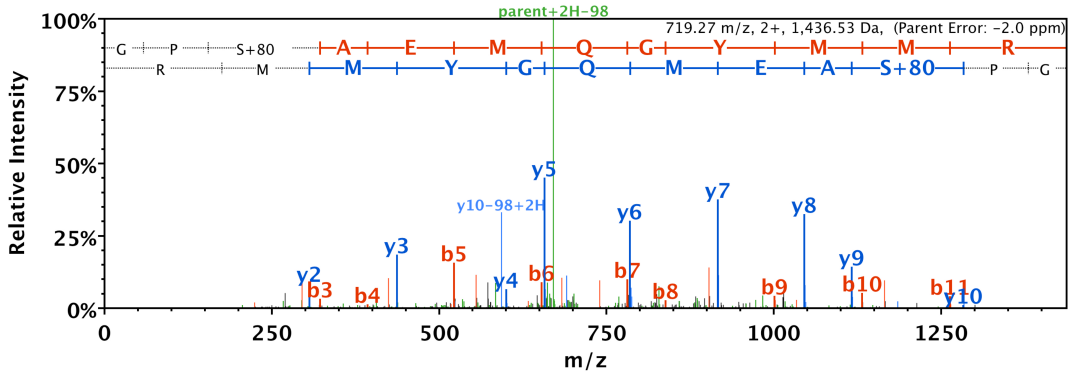
This section contains mass spectra collected for Chapters 3 and 4. Collected MS2 and MS3 spectra are shown for each phosphorylation site. When several unique peptides were detected for phosphorylation at a single residue, we only show the mass spectra for the highest scoring peptide (Mascot score). When a peptide phosphorylated at multiple residues was detected, we show all of its associated MS2 spectra.

Appendix 1:
Mass Spectra for CamKII, CDK5 and PLK2 Phosphorylation Sites in synGAP

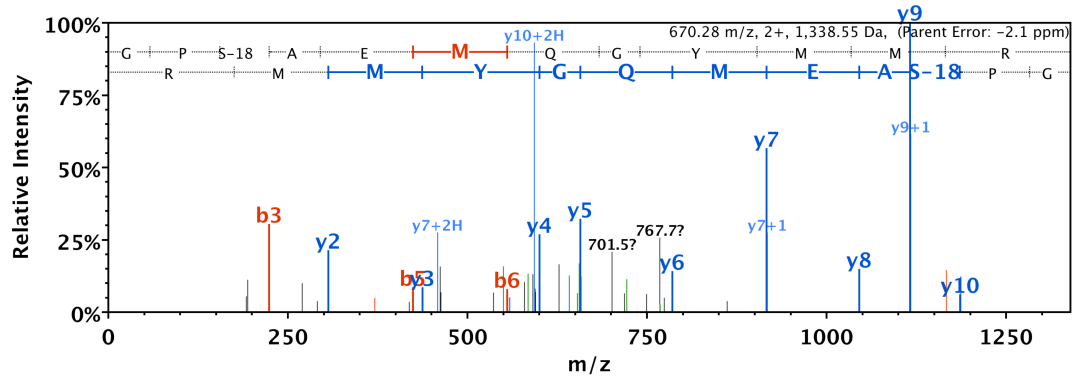
MASS SPECTRA FOR CAMKII PHOSPHORYLATION SITES IN SYNGAP



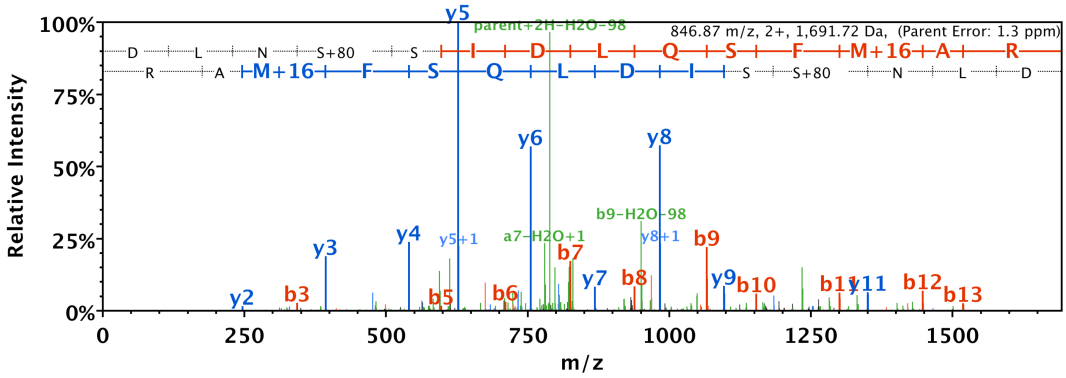
MS2 **S737**


parent+2H-98
719.27 m/z, 2+, 1,436.53 Da, (Parent Error: -2.0 ppm)
G P S+80 A E M Q G Y M M R
R M M Y G Q M E A S+80 P G
Relative Intensity
100%
75%
50%
25%
0%
y5
y10-98+2H
y6
y7
y8
y3
b5
y9
y2
b3
b4
y4
b6
b7
b8
b9
b10
b11
y10
0
250
500
750
1000
1250
m/z

MS3


y9
y10+2H
670.28 m/z, 2+, 1,338.55 Da, (Parent Error: -2.1 ppm)
G P S-18 A E M Q G Y M M R
R M M Y G Q M E A S-18 P G
Relative Intensity
100%
75%
50%
25%
0%
y9+1
y7
b3
y7+2H
y5
y4
767.7?
701.5?
y7+1
y2
y6
y8
b5
b6
y10
0
250
500
750
1000
1250
m/z

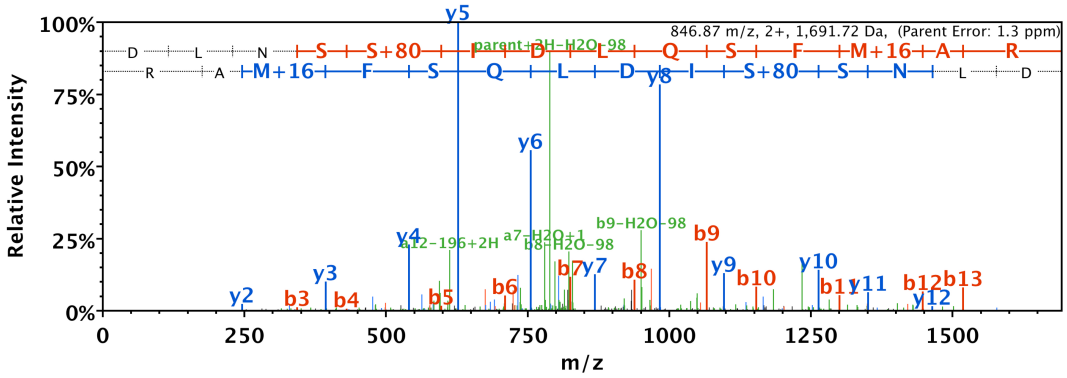
MS2 **S750**



MS3

No MS3 Spectra

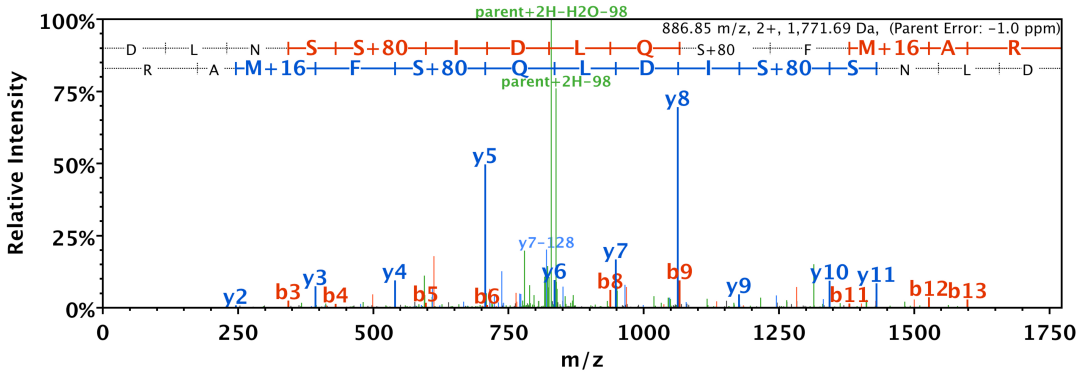
MS2 **S751**



MS3

No MS3 Spectra

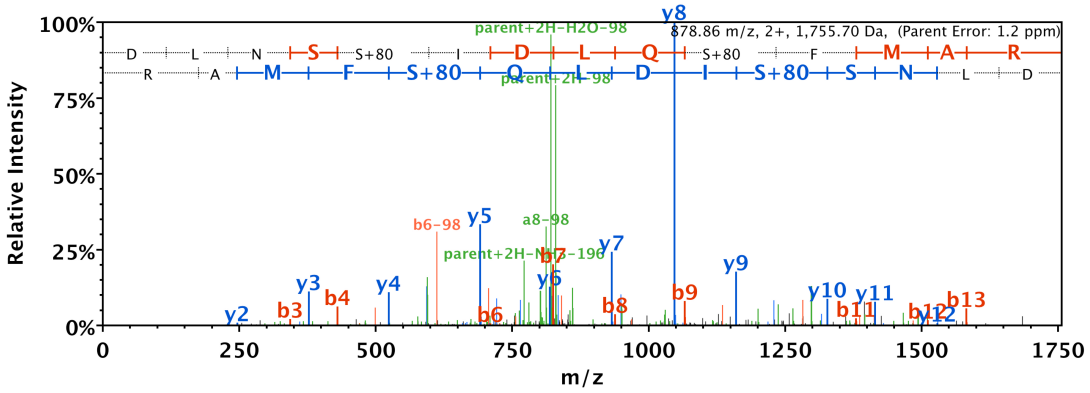
MS2 **S756**



MS3

No MS3 Spectra

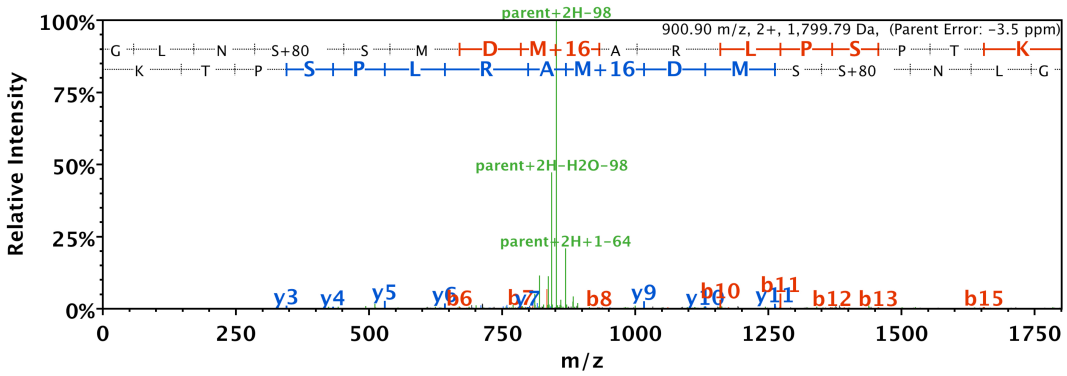
MS2 **S751 & S756**



MS3

No MS3 Spectra

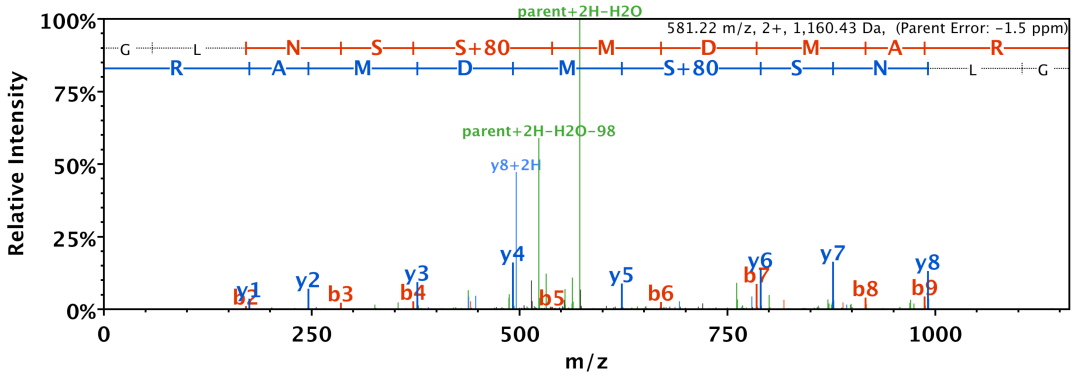
MS2 **S764**



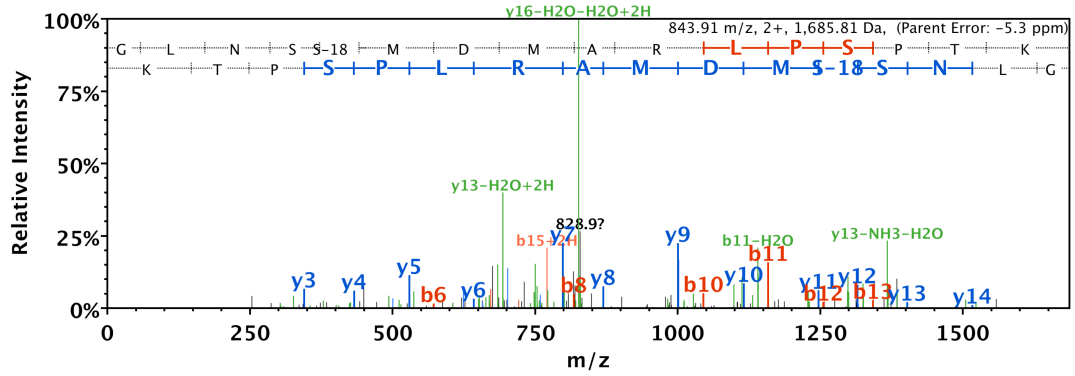
MS3

No MS3 Spectra

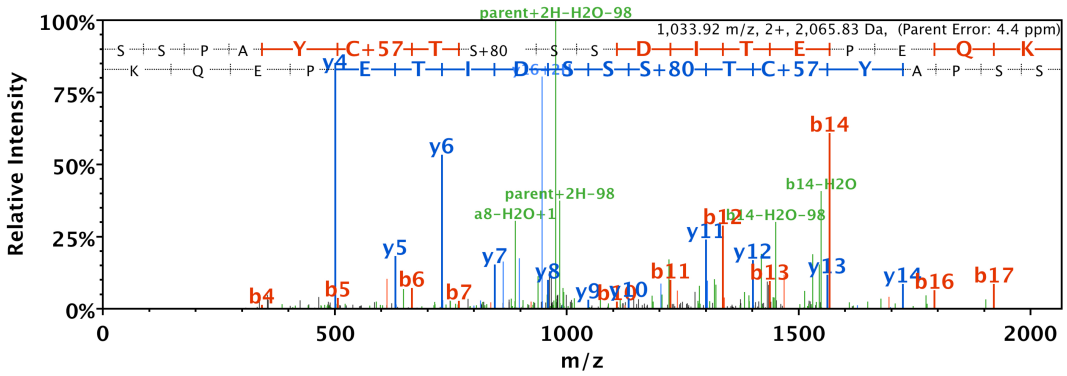
MS2 **S765**


parent+2H-H2O
581.22 m/z, 2+, 1,160.43 Da, (Parent Error: -1.5 ppm)
G L N S S+80 M D M A R
R A M D M S+80 S N L G
parent+2H-H2O-98
y8+2H
Relative Intensity
100%
75%
50%
25%
0%
0
250
500
750
1000
m/z

MS3


y16-H2O-H2O+2H
843.91 m/z, 2+, 1,685.81 Da, (Parent Error: -5.3 ppm)
G L N S S-18 M D M A R L P S P T K
K T P S P L R A M D M S-18 S N L G
y13-H2O+2H
828.9?
b15-H2O
b11-H2O
y13-NH3-H2O
Relative Intensity
100%
75%
50%
25%
0%
0
250
500
750
1000
1250
1500
m/z

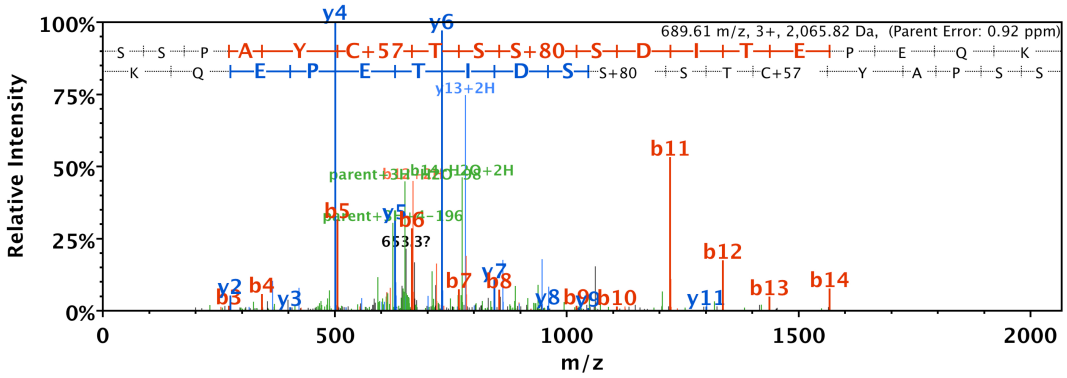
MS2 **S808**



MS3

No MS3 Spectra

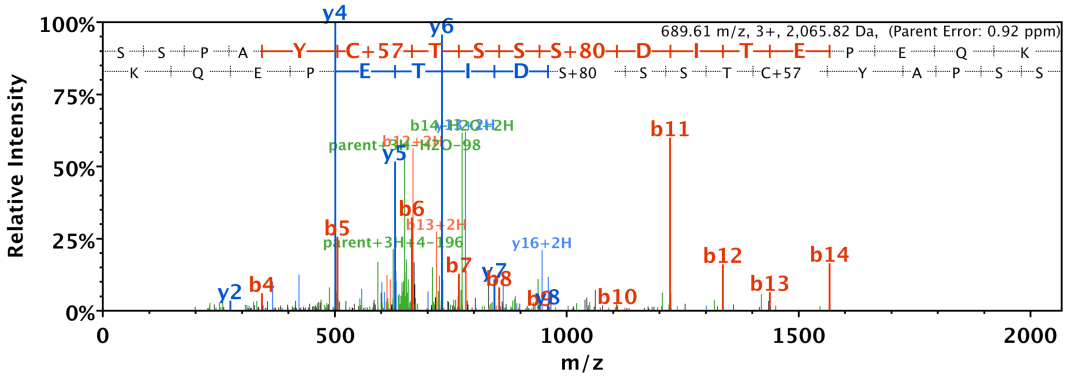
MS2 **S809**



MS3

No MS3 Spectra

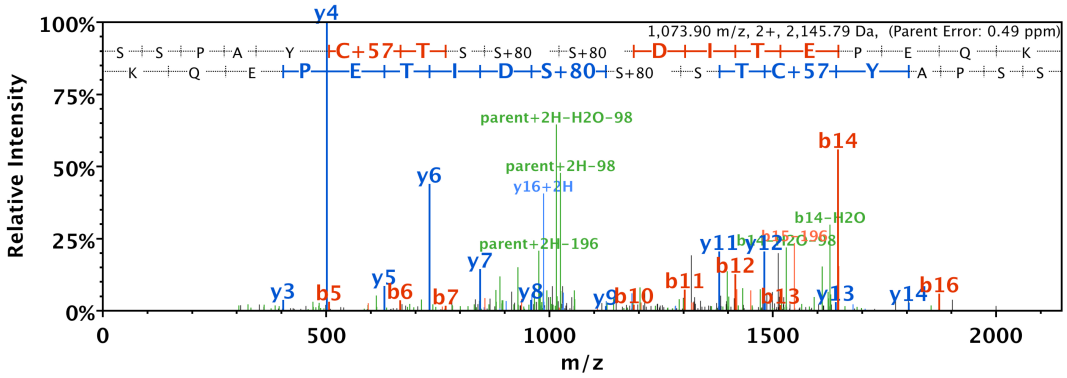
MS2 **S810**



MS3

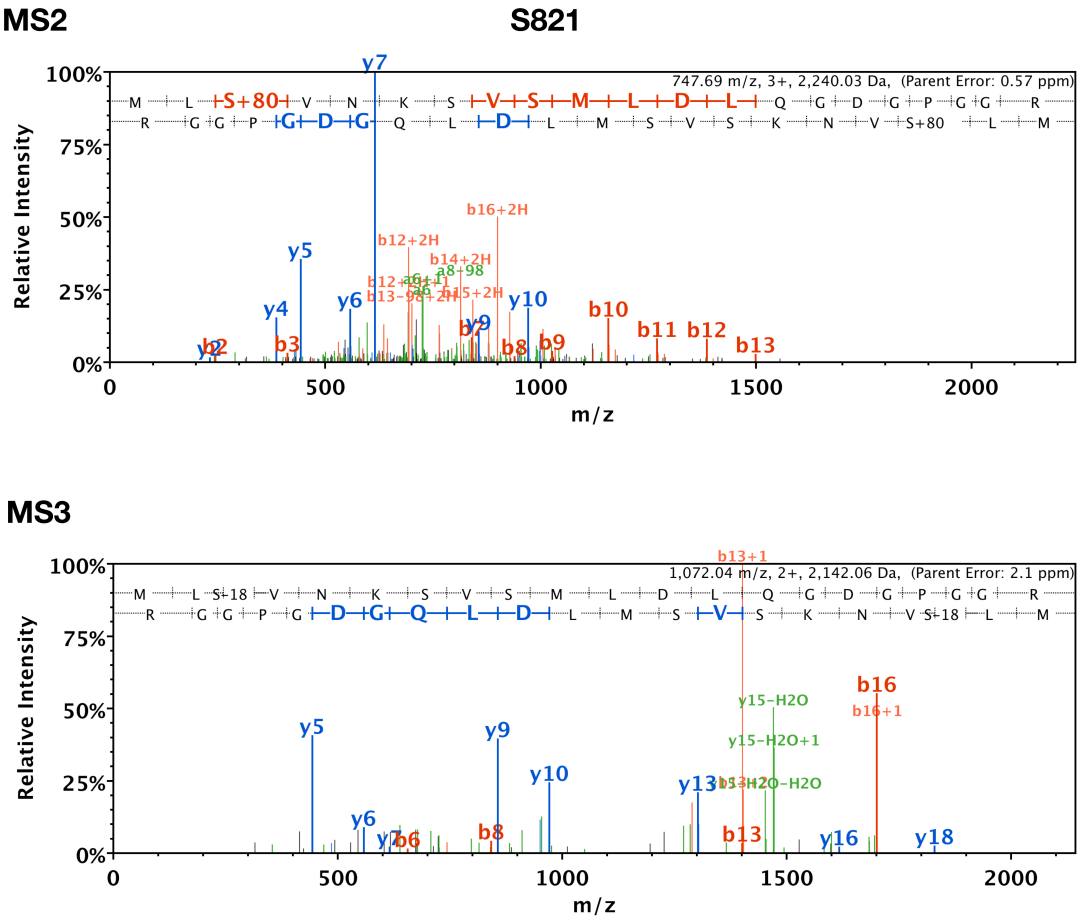
No MS3 Spectra

MS2 **S809 & S810**

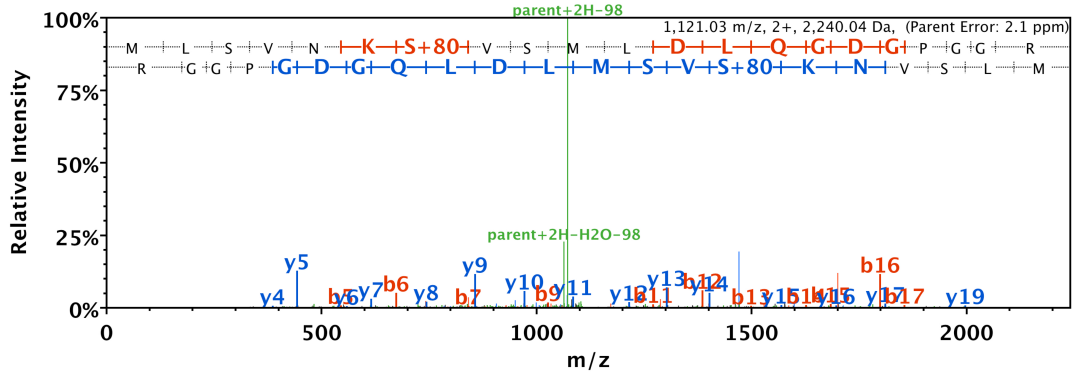


MS3

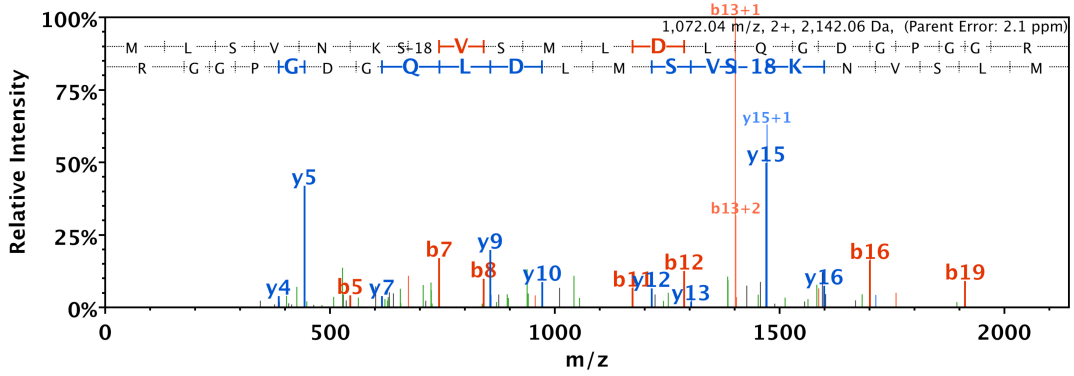
No MS3 Spectra


MS2
S821
747.69 m/z, 3+, 2,240.03 Da, (Parent Error: 0.57 ppm)
M L S+80 V N K S V S M L D L Q G D G P G G R
R G G P G D G Q L D L M S V S K N V S+80 L M
Relative Intensity
m/z
y7
y5
y4
y6
y10
b16+2H
b12+2H
b14+2H
a8+98
b10
b11
b12
b13
b2
b3
b8
b9
y2
y9
MS3
1,072.04 m/z, 2+, 2,142.06 Da, (Parent Error: 2.1 ppm)
M L S+18 V N K S V S M L D L Q G D G P G G R
R G G P G D G Q L D L M S V S K N V S+18 L M
b13+1
b16
b16+1
y15-H2O
y15-H2O+1
y13-H2O-H2O
y5
y9
y10
y13
y6
b8
b13
y16
y18

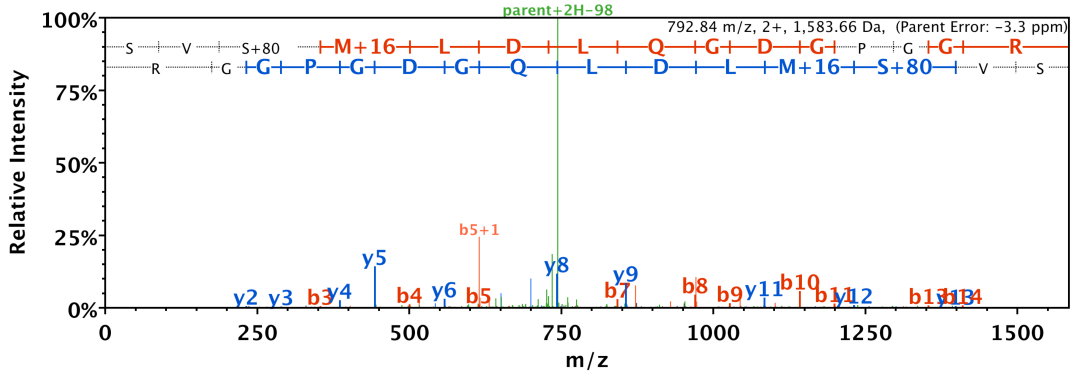
MS2 **S825**


parent+2H-98
1,121.03 m/z, 2+, 2,240.04 Da, (Parent Error: 2.1 ppm)
M L S V N K S+80 V S M L D L Q G D G P G G R
R G G P G D G Q L D L M S V S+80 K N V S L M
parent+2H-H2O-98
Relative Intensity
100%
75%
50%
25%
0%
0
500
1000
1500
2000
m/z
y5
y4
y9
b16
y19

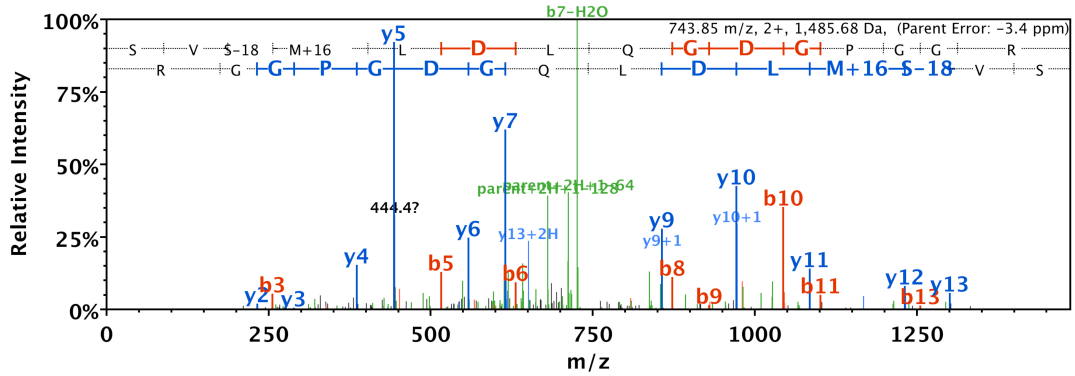
MS3


b13+1
1,072.04 m/z, 2+, 2,142.06 Da, (Parent Error: 2.1 ppm)
M L S V N K S-18 V S M L D L Q G D G P G G R
R G G P G D G Q L D L M S V S-18 K N V S L M
y15+1
y15
b13+2
Relative Intensity
100%
75%
50%
25%
0%
0
500
1000
1500
2000
m/z
y5
y4
b7
y9
b8
y10
b12
y13
b16
y16
b19

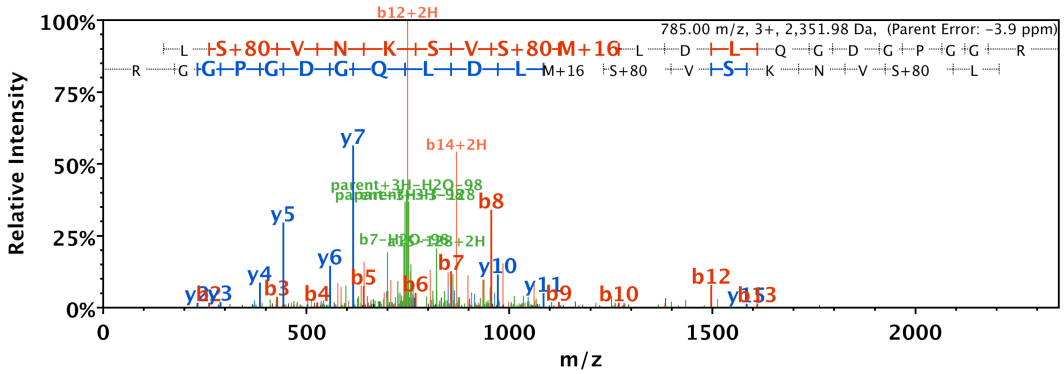
MS2 **S827**


parent+2H-98
792.84 m/z, 2+, 1,583.66 Da, (Parent Error: -3.3 ppm)
S V S+80 M+16 L D L Q G D G P G G R
R G G P G D G Q L D L M+16 S+80 V S
Relative Intensity
100%
75%
50%
25%
0%
b5+1
y5
y8
y9
y2 y3 b3 y4 b4 y6 b5 b7 b8 b9 y11 b10 y12 b12 b13 y13
0
250
500
750
1000
1250
1500
m/z

MS3


b7-H2O
743.85 m/z, 2+, 1,485.68 Da, (Parent Error: -3.4 ppm)
S V S-18 M+16 L D L Q G D G P G G R
R G G P G D G Q L D L M+16 S-18 V S
Relative Intensity
100%
75%
50%
25%
0%
y5
y7
444.4?
parent+2H+1-128
parent+2H-164
y10
y10+1
b10
y9
y9+1
y6
y13+2H
y4
b5
b6
b8
y11
b11
y12
y13
b3
y2
y3
b9
b13
0
250
500
750
1000
1250
m/z

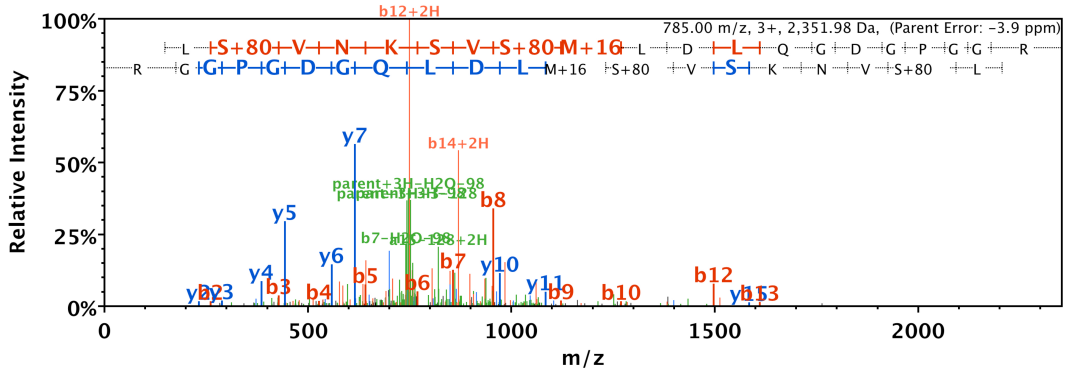
MS2 **S821 & S825**



MS3

No MS3 Spectra

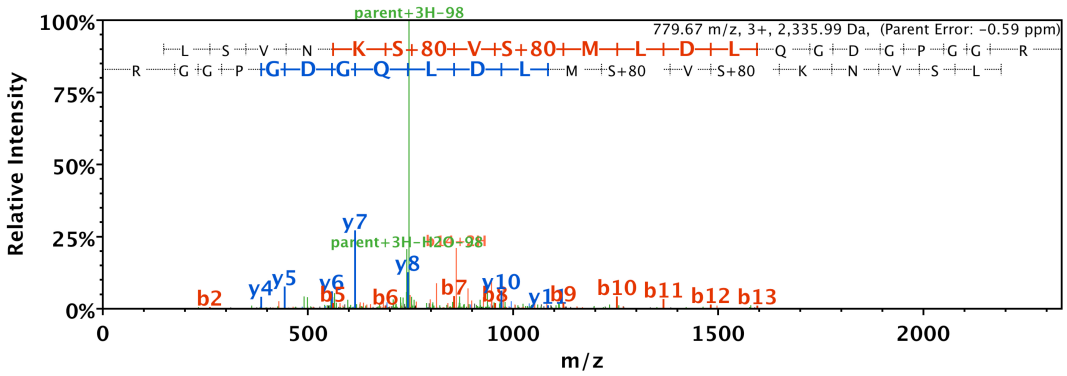
MS2 **S821 & S827**



MS3

No MS3 Spectra

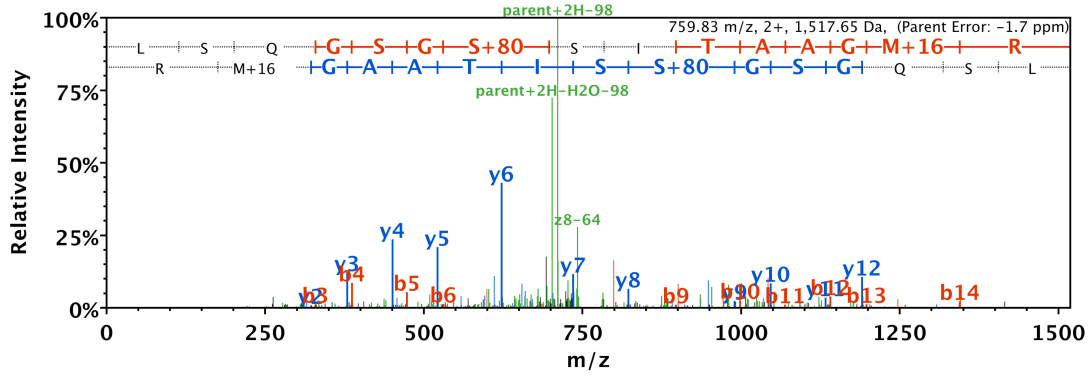
MS2 **S825 & S827**



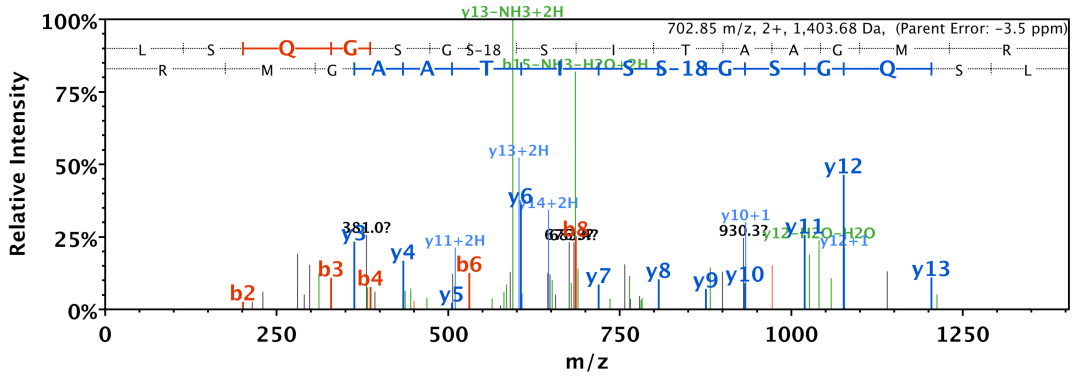
MS3

No MS3 Spectra

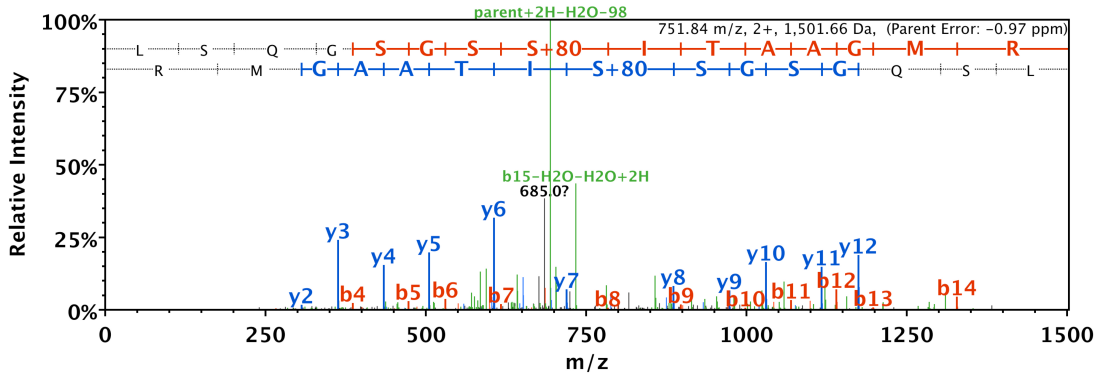
MS2 **S882**


parent+2H-98
759.83 m/z, 2+, 1,517.65 Da, (Parent Error: -1.7 ppm)
L S Q G S G S+80 S I T A A G M+16 R
R M+16 G A A T I S S+80 G S G Q S L
parent+2H-H2O-98
Relative Intensity
100%
75%
50%
25%
0%
y6
z8-64
y4
y5
y3
b4
b5
b2
b6
y7
y8
b9
y10
y12
b10
b11
b12
b13
b14
0
250
500
750
1000
1250
1500
m/z

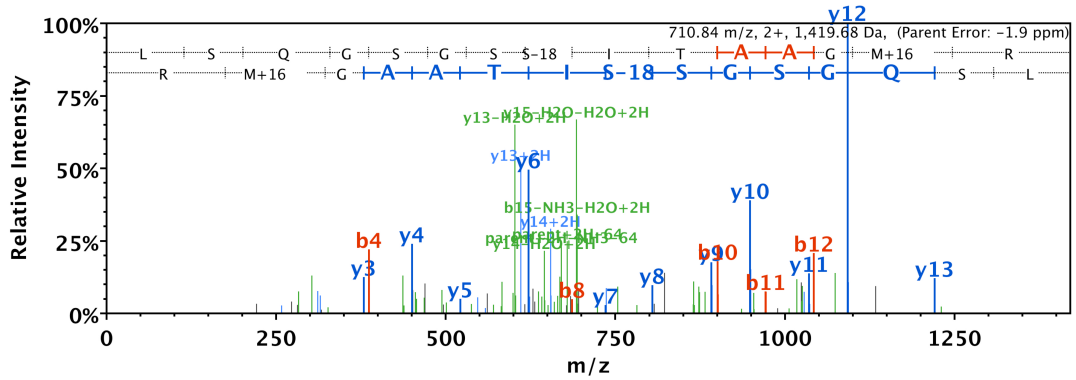
MS3


y13-NH3+2H
702.85 m/z, 2+, 1,403.68 Da, (Parent Error: -3.5 ppm)
L S Q G S G S-18 S I T A A G M R
R M G A A T I S S-18 G S G Q S L
b15-NH3-H2O+2H
Relative Intensity
100%
75%
50%
25%
0%
y13+2H
y12
y6
y14+2H
y10+1
381.0?
y3
y11+2H
y4
b3
b4
b6
b2
y5
y7
y8
y9
y10
y13
0
250
500
750
1000
1250
m/z

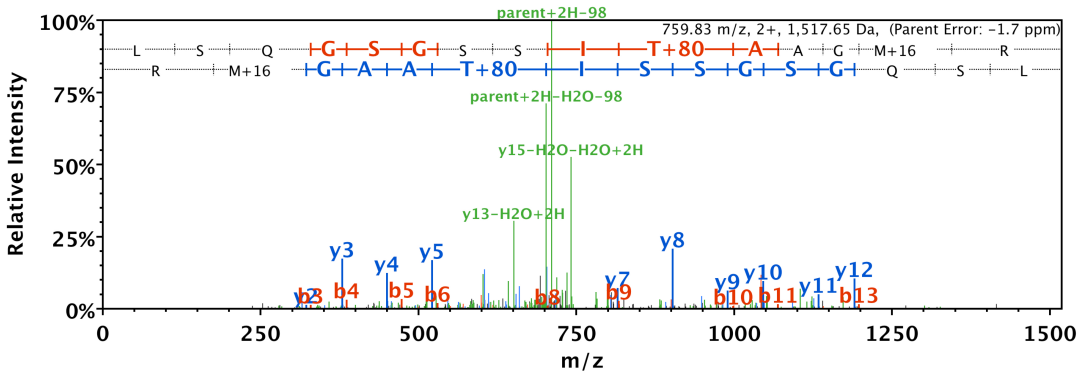
MS2 **S883**


parent+2H-H2O-98
751.84 m/z, 2+, 1,501.66 Da, (Parent Error: -0.97 ppm)
L S Q G S G S S+80 I T A A G M R
R M G A A T I S+80 S G S G Q S L
b15-H2O-H2O+2H
685.0?
Relative Intensity
m/z

MS3


710.84 m/z, 2+, 1,419.68 Da, (Parent Error: -1.9 ppm)
L S Q G S G S S-18 I T A A G M+16 R
R M+16 G A A T I S-18 S G S G Q S L
b15-NH3-H2O+2H
Relative Intensity
m/z

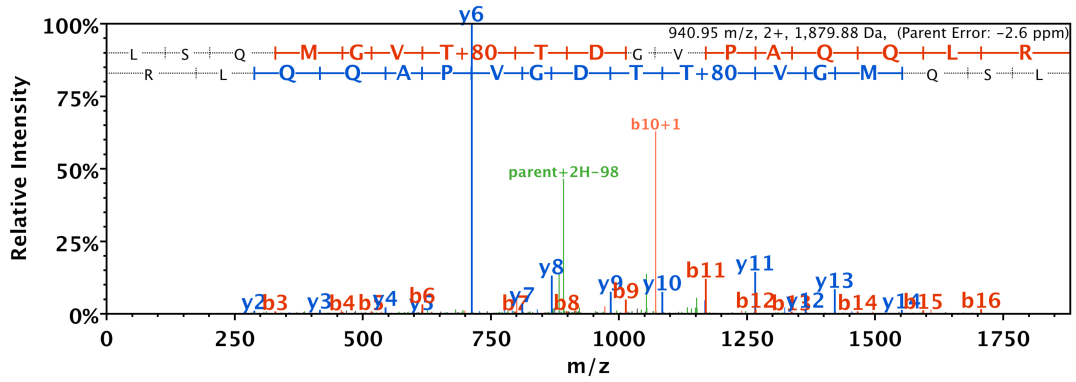
MS2 **T885**



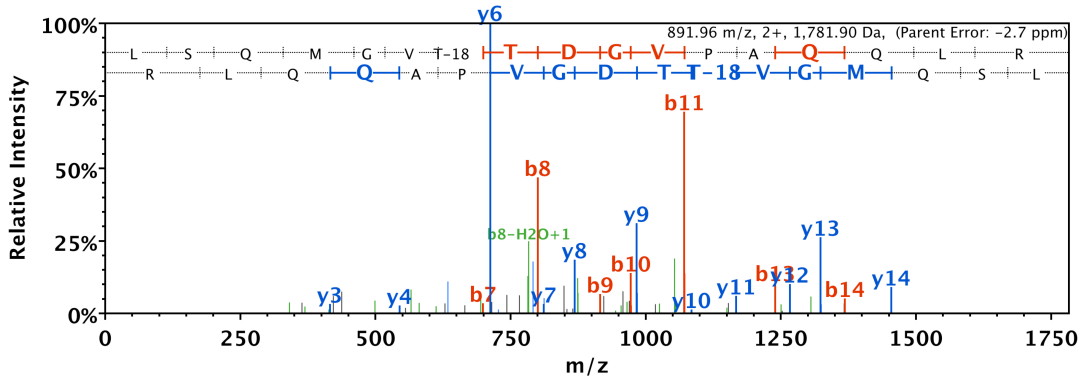
MS3

No MS3 Spectra

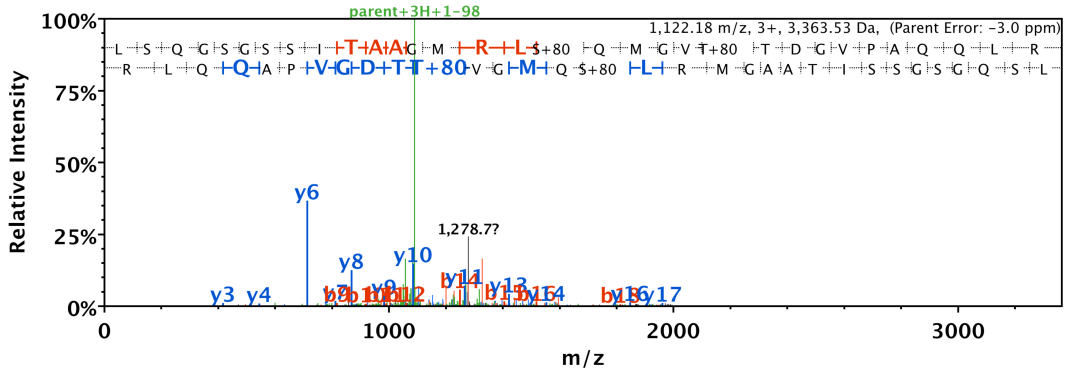
MS2 **T897**


940.95 m/z, 2+, 1,879.88 Da, (Parent Error: -2.6 ppm)
L S Q M G V T+80 T D G V P A Q Q L R
R L Q Q A P V G D T T+80 V G M Q S L
y6
b10+1
parent+2H-98
y8
b11
y11
y13
y9
y10
b9
b8
y7
y2
b3
y3
b4
b5
y4
b6
y5
b12
y12
b14
y14
b15
b16
Relative Intensity
100%
75%
50%
25%
0%
0
250
500
750
1000
1250
1500
1750
m/z

MS3


891.96 m/z, 2+, 1,781.90 Da, (Parent Error: -2.7 ppm)
L S Q M G V T-18 T D G V P A Q Q L R
R L Q Q A P V G D T T-18 V G M Q S L
y6
b11
b8
y9
y13
b8-H2O+1
y8
b10
b9
y3
y4
b7
y7
y10
y11
b12
b14
y14
Relative Intensity
100%
75%
50%
25%
0%
0
250
500
750
1000
1250
1500
1750
m/z

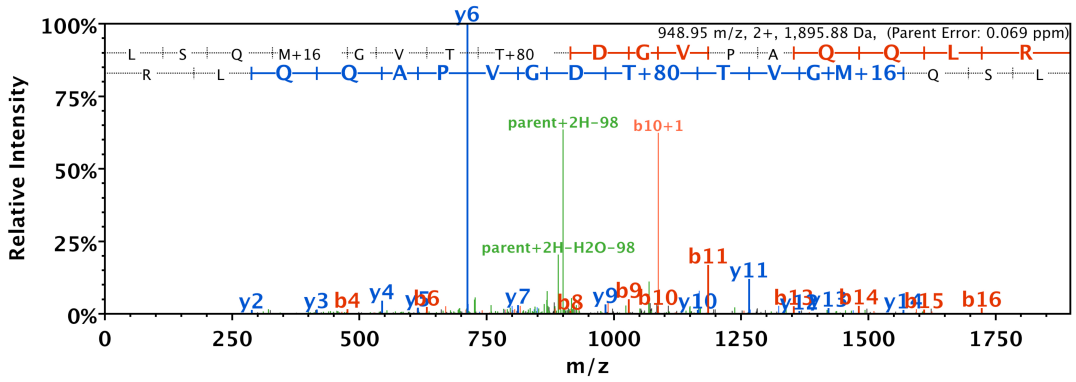
MS2 **S892 & T897**



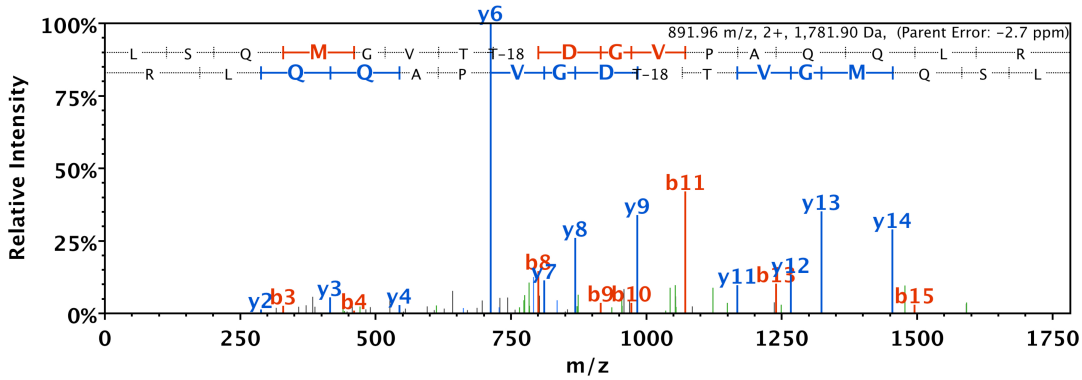
MS3

No MS3 Spectra

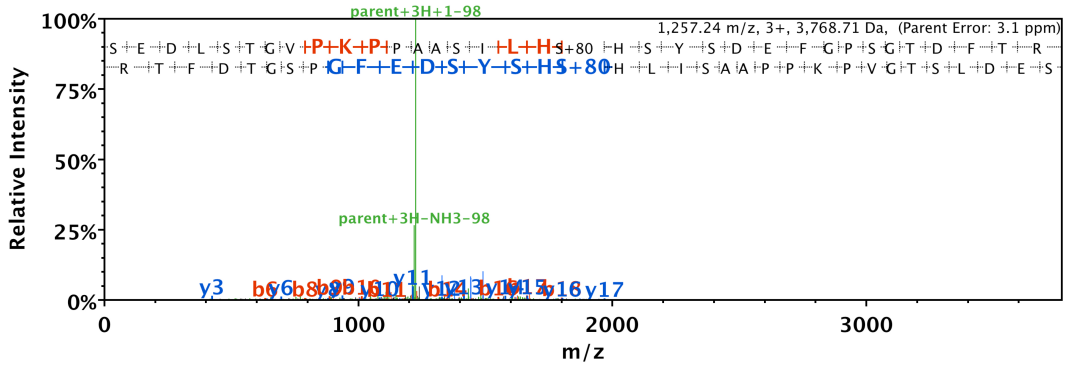
MS2 **T898**


948.95 m/z, 2+, 1,895.88 Da, (Parent Error: 0.069 ppm)
L S Q M+16 G V T T+80 D G V P A Q Q L R
R L Q Q A P V G D T+80 T V G M+16 Q S L
parent+2H-98
b10+1
parent+2H-H2O-98
Relative Intensity
m/z

MS3


891.96 m/z, 2+, 1,781.90 Da, (Parent Error: -2.7 ppm)
L S Q M G V T T-18 D G V P A Q Q L R
R L Q Q A P V G D T-18 T V G M Q S L
Relative Intensity
m/z

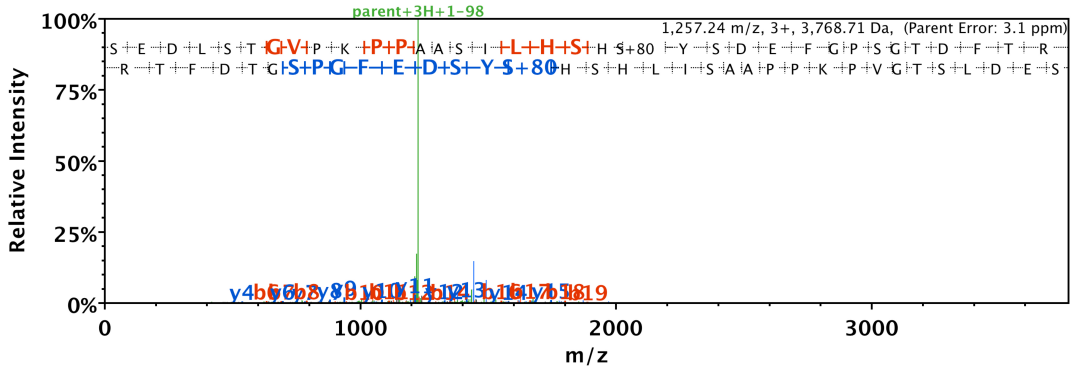
MS2 **S988**



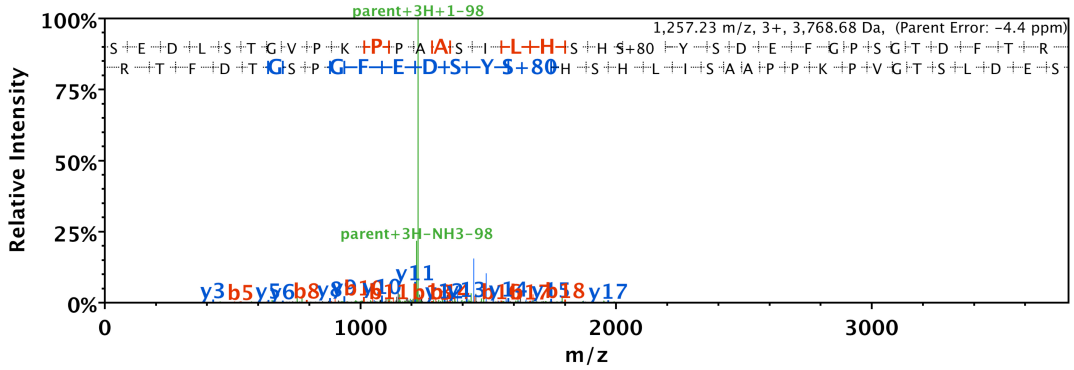
MS3

No MS3 Spectra

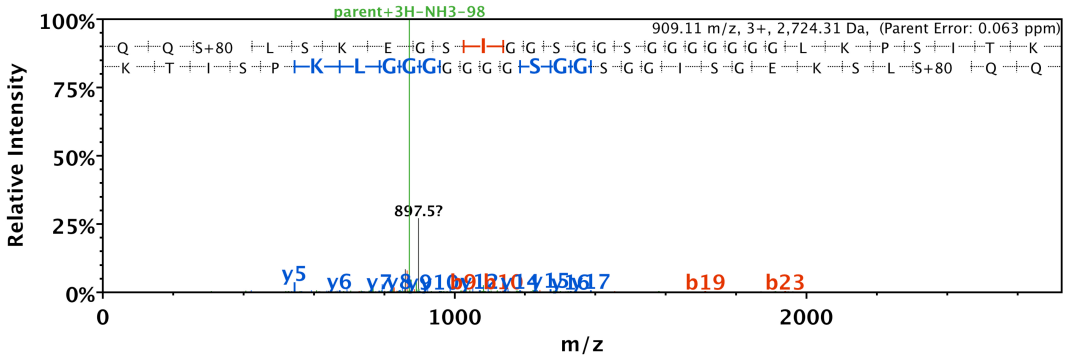
MS2 **S990**


parent+3H+1-98
1,257.24 m/z, 3+, 3,768.71 Da, (Parent Error: 3.1 ppm)
S E D L S T G V P K P P A A S I L H S+80 Y S D E F G P S G T D F T R
R T F D T G S P G F E D S Y S+80 H S H L I S A A P P K P V G T S L D E S
Relative Intensity
100%
75%
50%
25%
0%
0
1000
2000
3000
m/z

MS3


parent+3H+1-98
1,257.23 m/z, 3+, 3,768.68 Da, (Parent Error: -4.4 ppm)
S E D L S T G V P K P P A A S I L H S+80 Y S D E F G P S G T D F T R
R T F D T G S P G F E D S Y S+80 H S H L I S A A P P K P V G T S L D E S
parent+3H-NH3-98
y11
y3 b5 y5 y6 b8 y8 y17
Relative Intensity
100%
75%
50%
25%
0%
0
1000
2000
3000
m/z

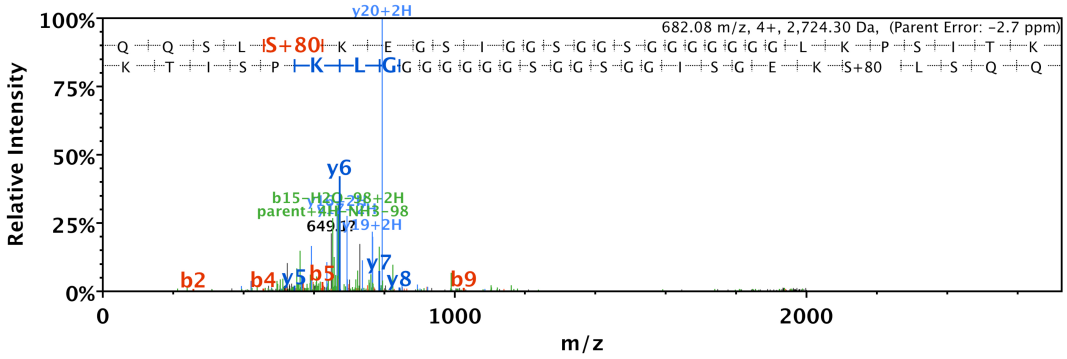
MS2 **S1093**



MS3

No MS3 Spectra

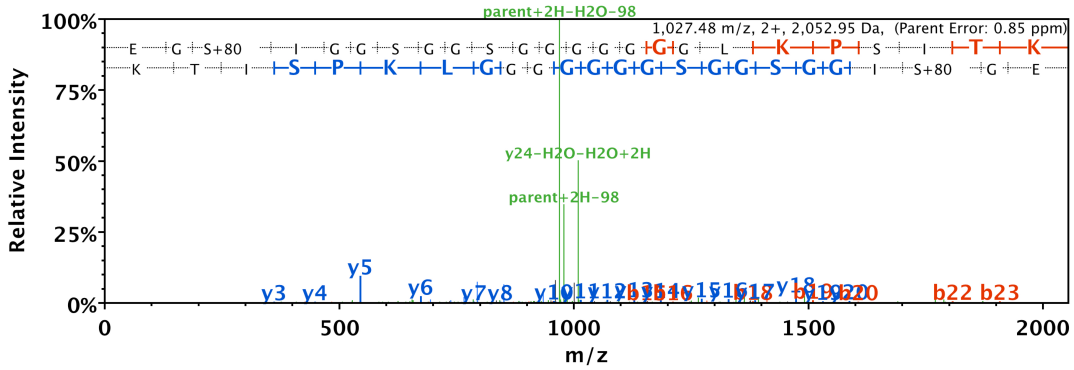
MS2 **S1095**



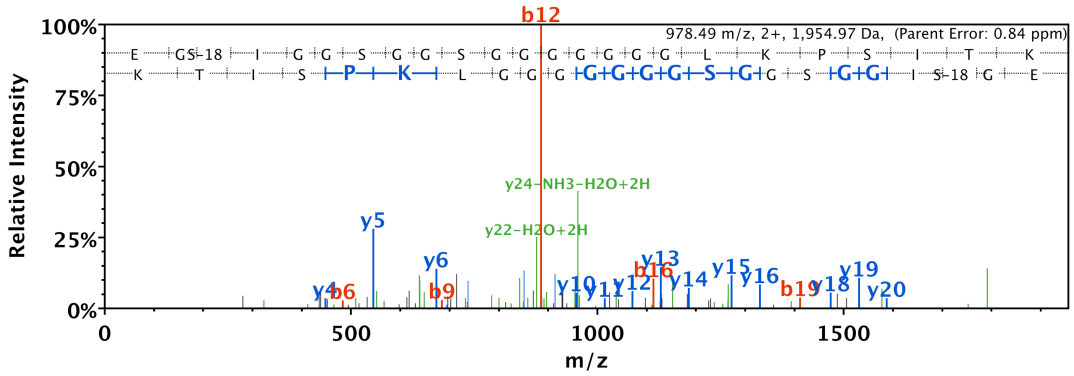
MS3

No MS3 Spectra

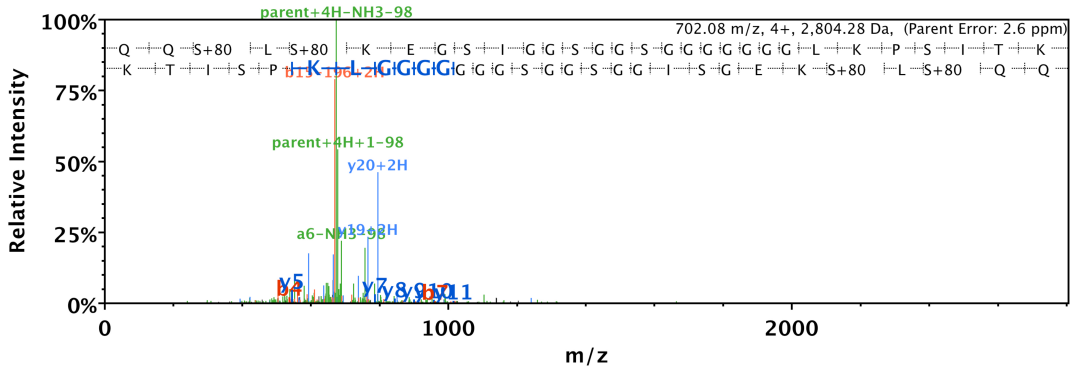
MS2 **S1099**



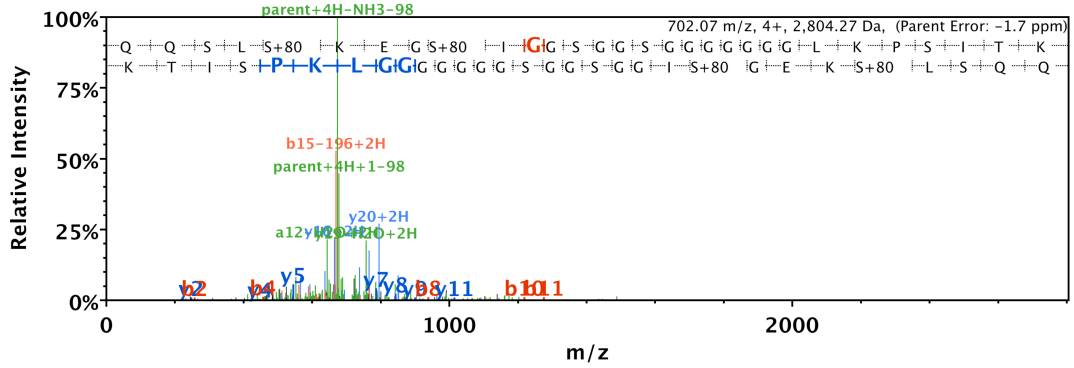
MS3



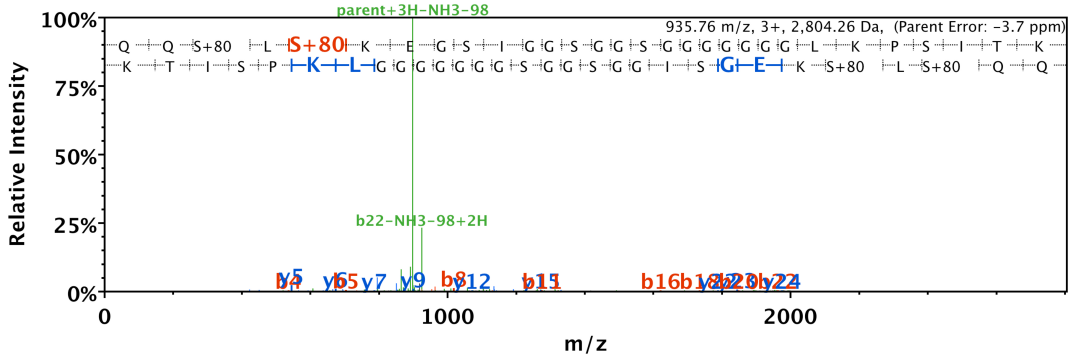
MS2 **S1093 & S1095**



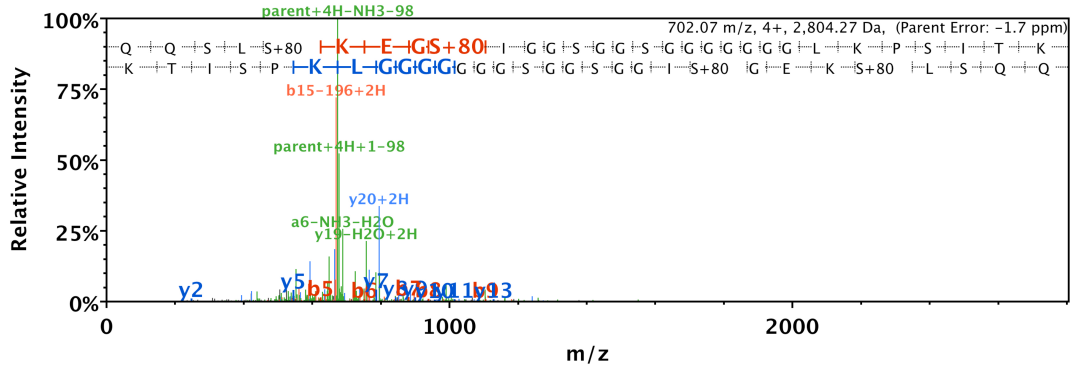
MS2



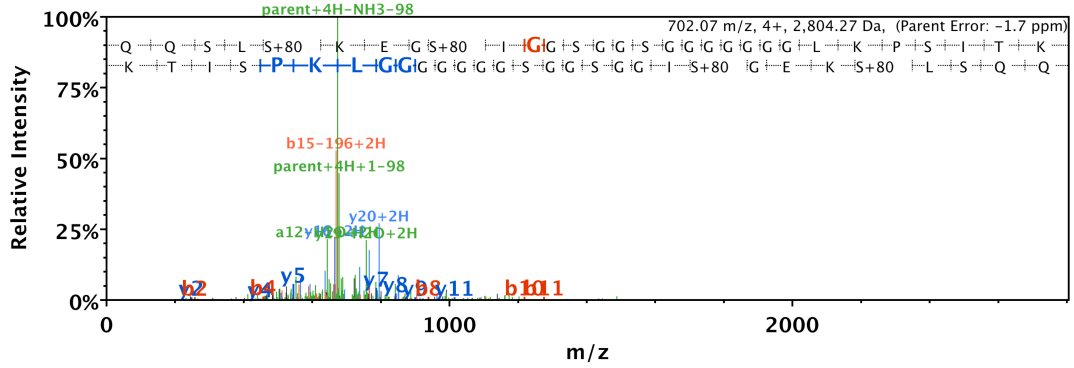
MS2



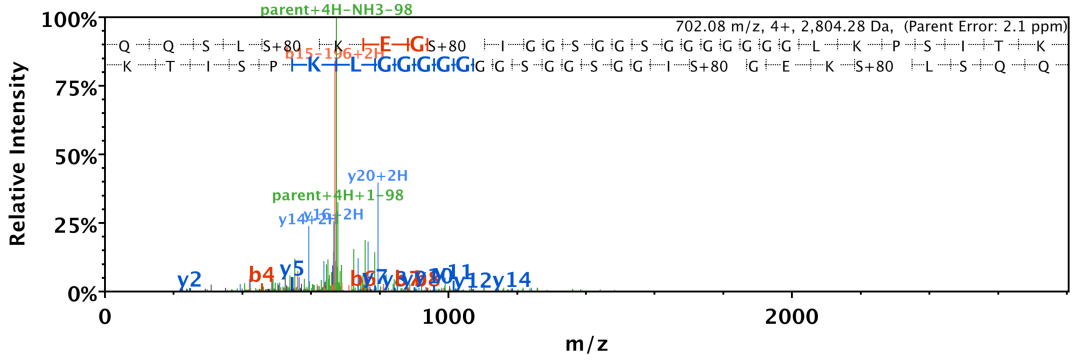
MS2 **S1095 & S1099**



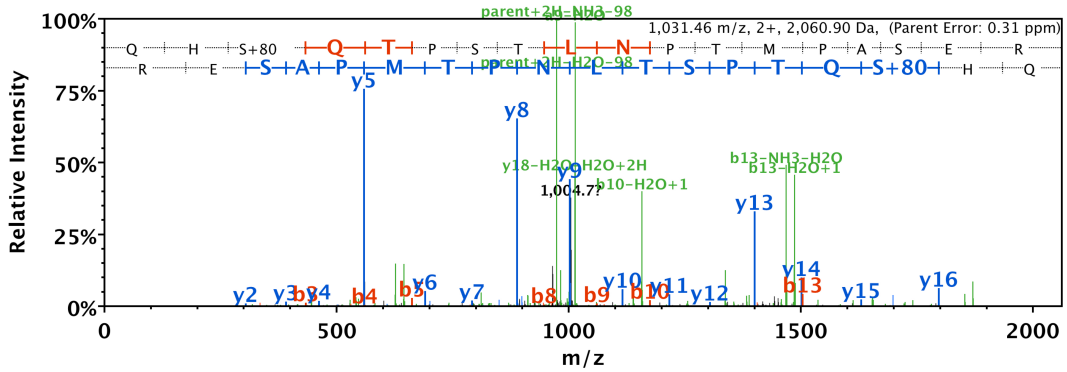
MS2



MS2



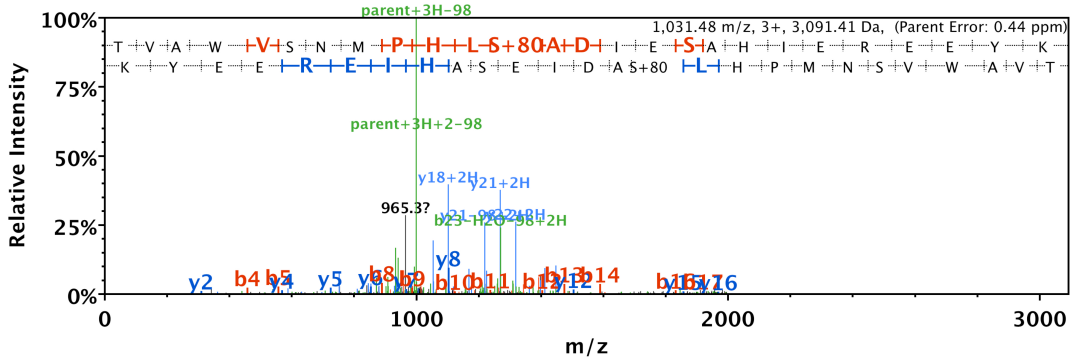
MS2 **S1123**



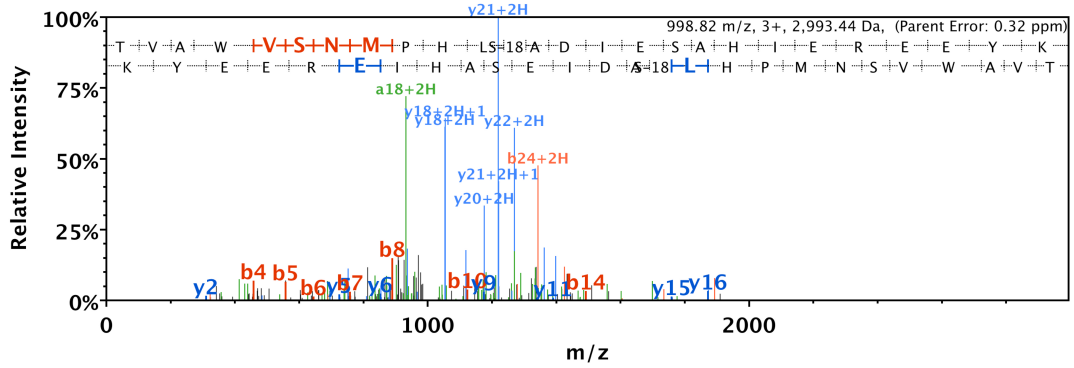
MS3

No MS3 Spectra

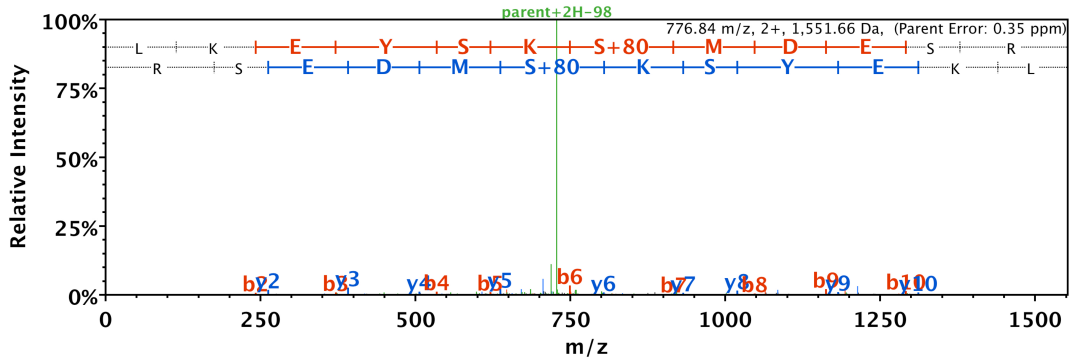
MS2 **S1150**



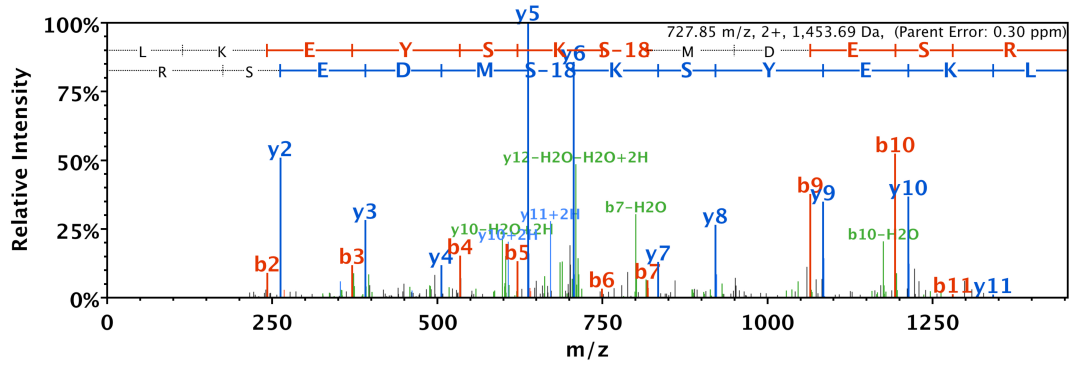
MS3



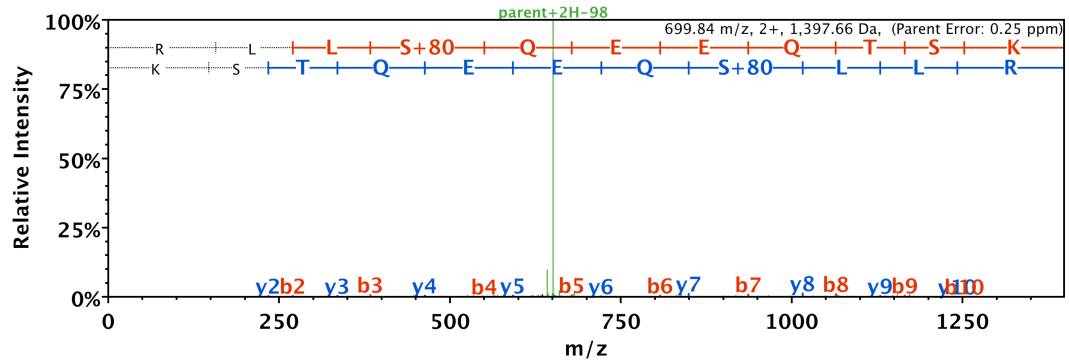
MS2 **S1171**



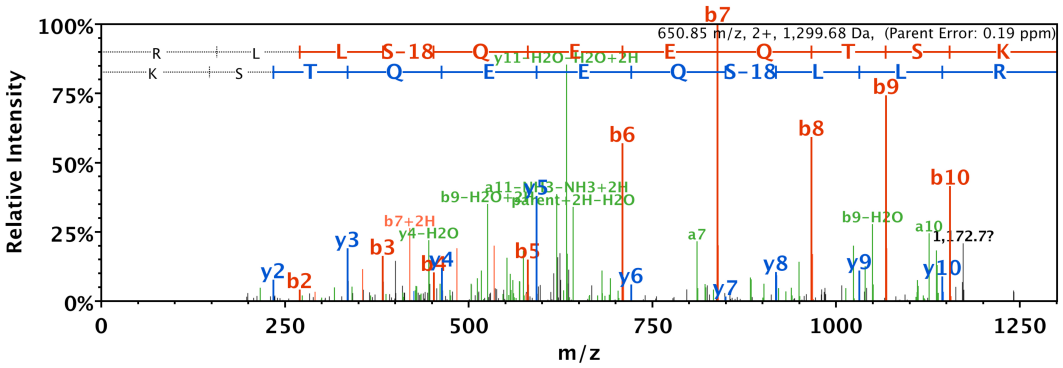
MS3



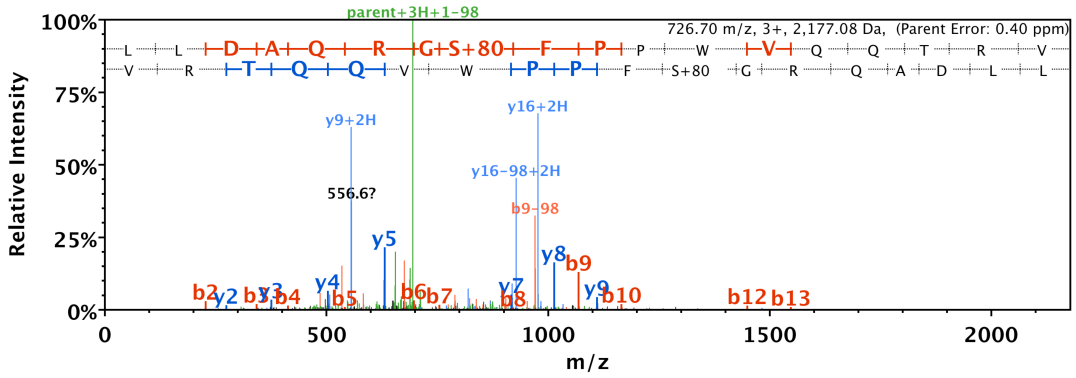
MS2 **S1210**



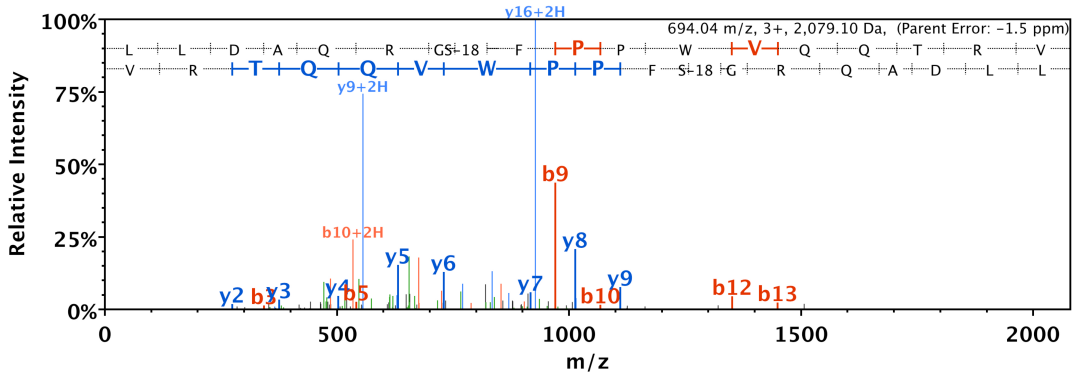
MS3



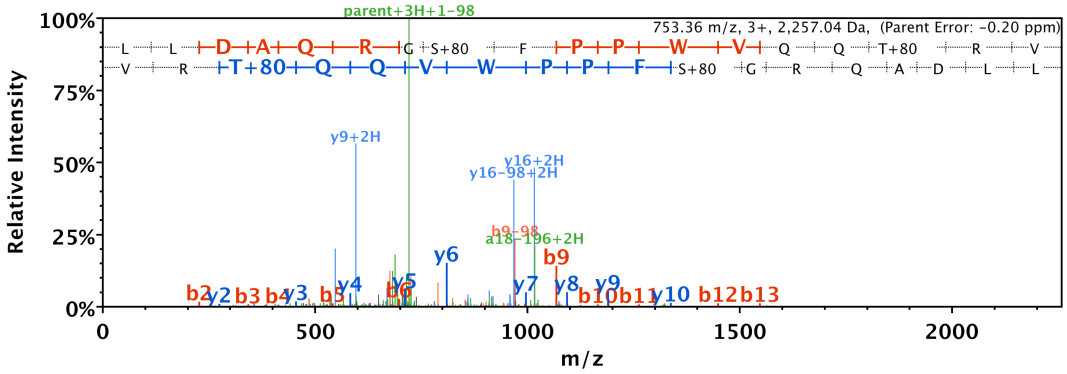
MS2 **S1283**



MS3



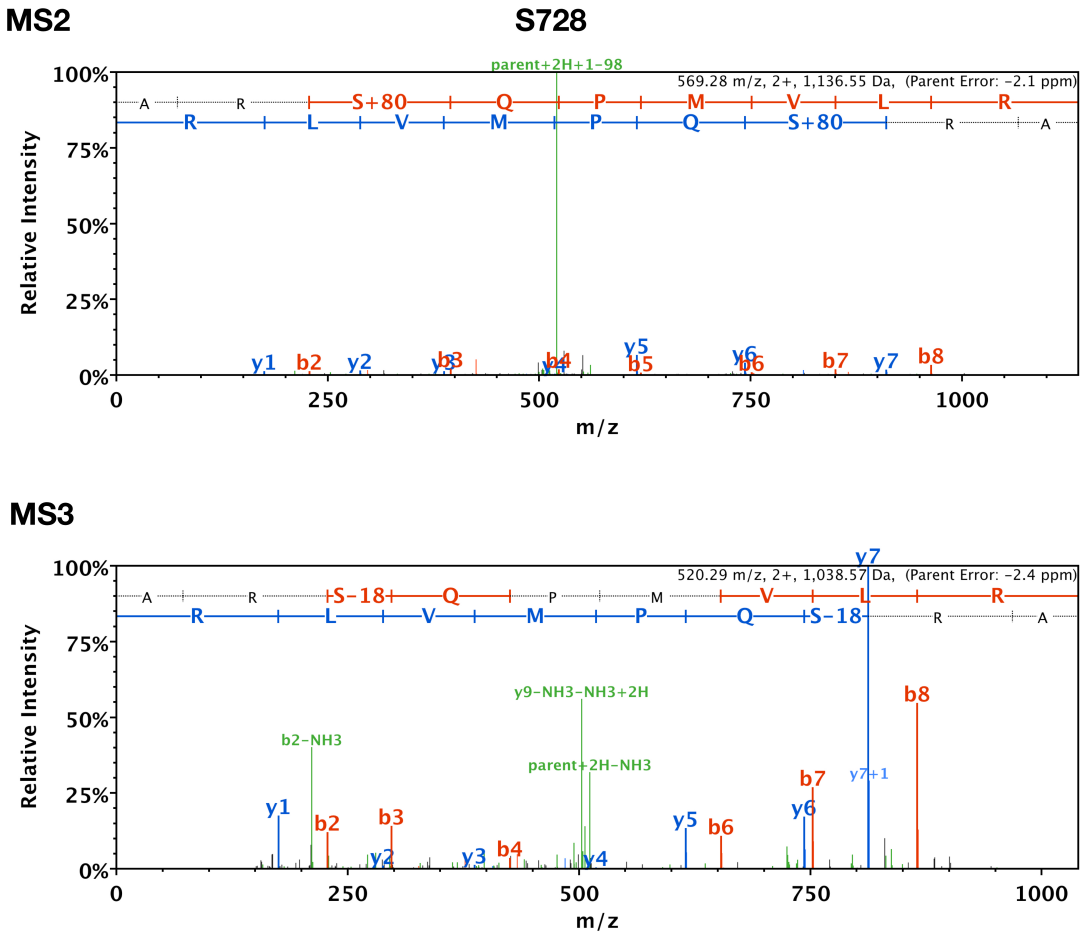
MS2 **S1283 & T1291**



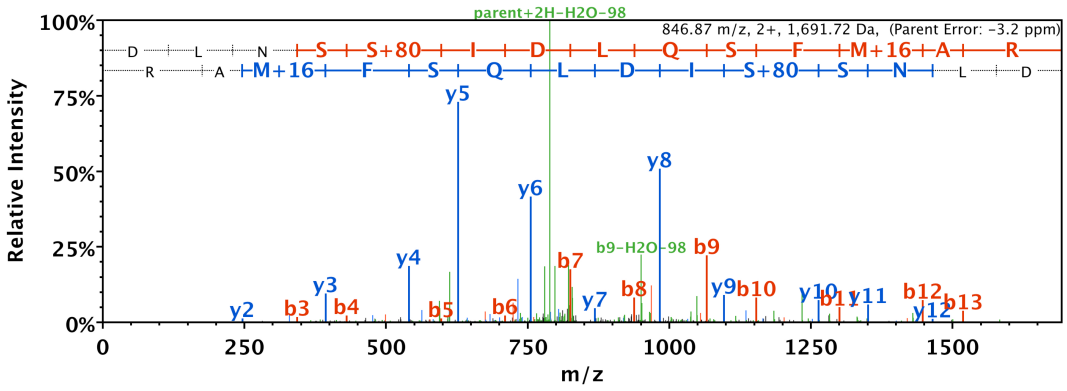
MS3

No MS3 Spectra

MASS SPECTRA FOR CDK5 PHOSPHORYLATION SITES IN SYNGAP



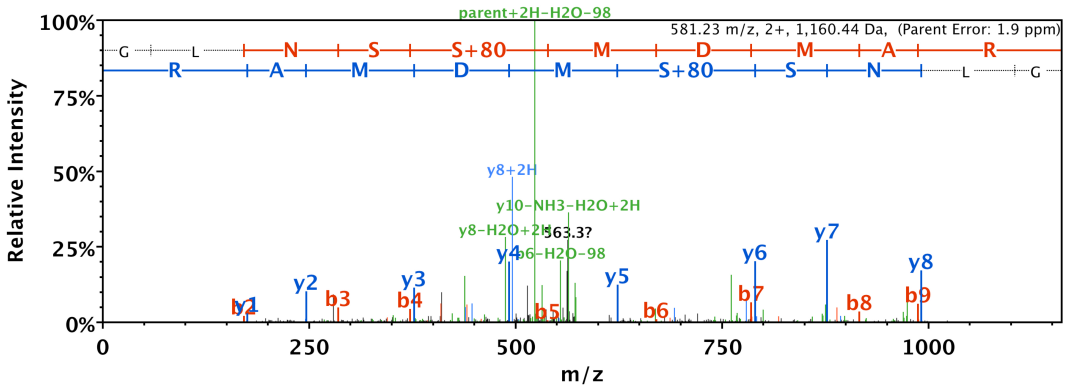
MS2 **S751**



MS3

No MS3 Spectra

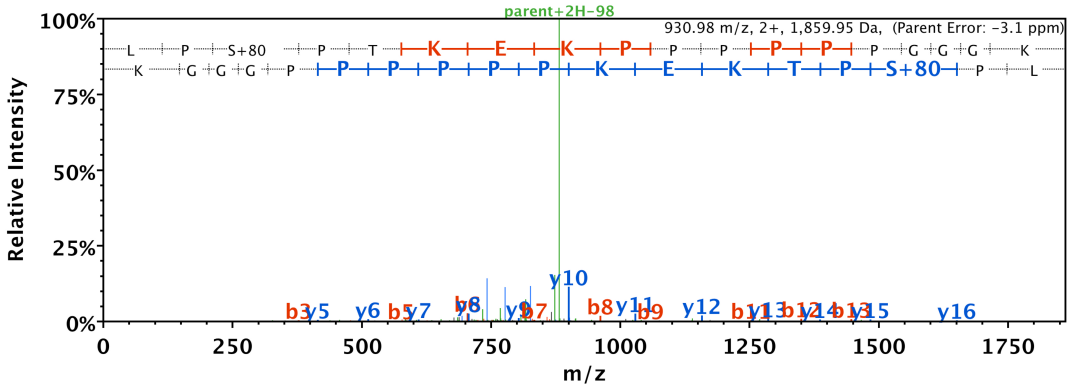
MS2 **S765**



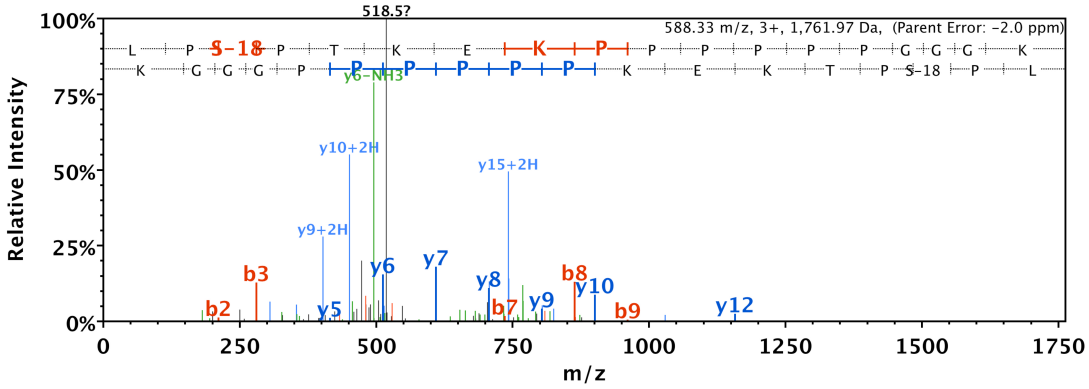
MS3

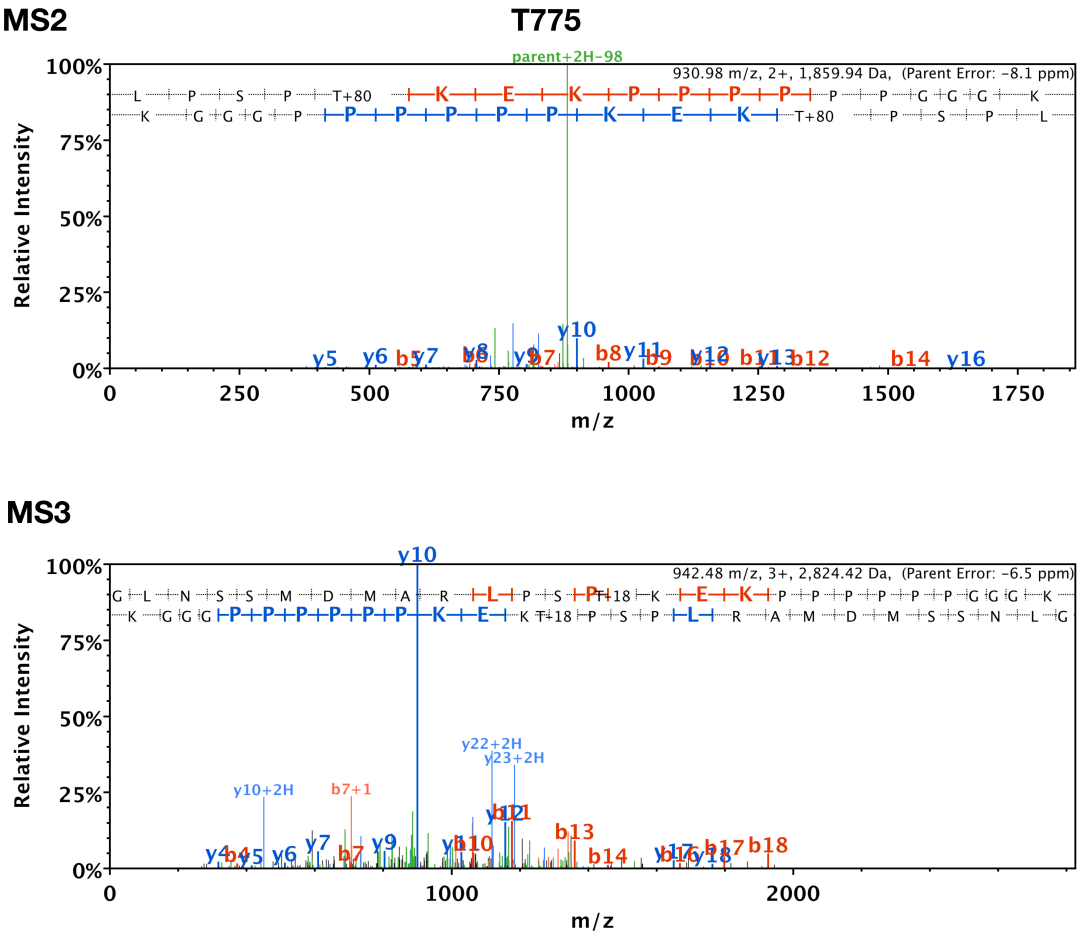
No MS3 Spectra

MS2 **S773**

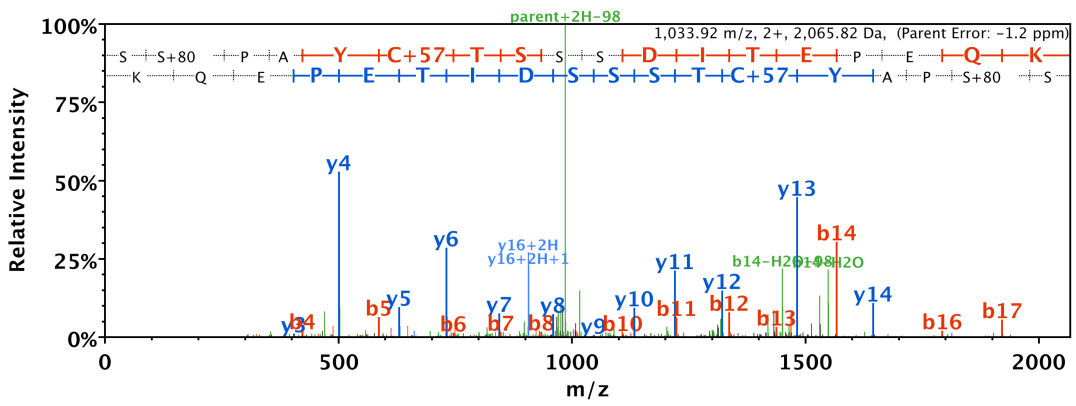

parent+2H-98
930.98 m/z, 2+, 1,859.95 Da, (Parent Error: -3.1 ppm)
L P S+80 P T K E K P P P P P P G G G K
K G G G P P P P P P K E K T P S+80 P L
Relative Intensity
100%
75%
50%
25%
0%
0
250
500
750
1000
1250
1500
1750
m/z
y10
b3 y5 y6 b5 y7 y8 y9 b7 b8 y11 b9 y12 y16

MS3

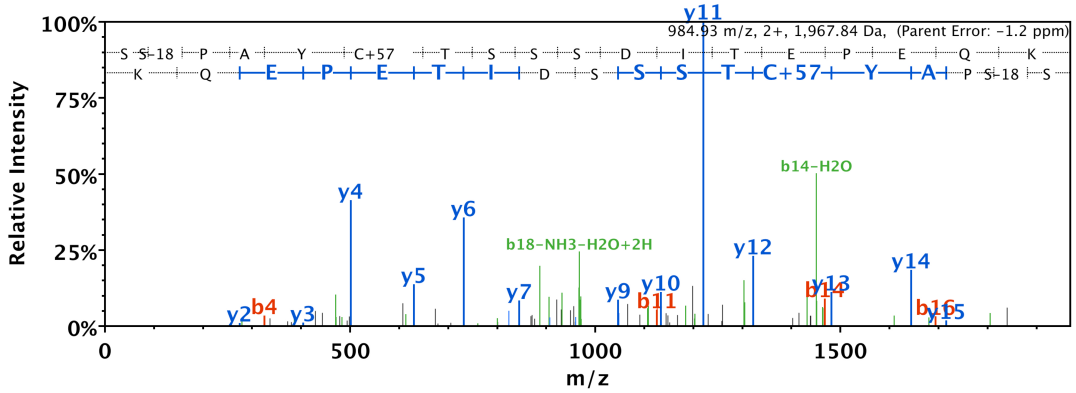

518.5?
588.33 m/z, 3+, 1,761.97 Da, (Parent Error: -2.0 ppm)
L P S-18 P T K E K P P P P P P G G G K
K G G G P P P P P P K E K T P S-18 P L
y6-NH3
Relative Intensity
100%
75%
50%
25%
0%
0
250
500
750
1000
1250
1500
1750
m/z
y10+2H
y15+2H
y9+2H
b3
b2
y5
y6
y7
y8
b7
y9
b8
y10
b9
y12


MS2
T775
parent+2H-98
930.98 m/z, 2+, 1,859.94 Da, (Parent Error: -8.1 ppm)
L P S P T+80 K E K P P P P P P G G G K
K G G G P P P P P P K E K T+80 P S P L
Relative Intensity
100%
75%
50%
25%
0%
y10
y5 y6 b7 y7 b8 y8 y9 b8 y11 b9 b10 y13 b11 b12 b14 y16
0 250 500 750 1000 1250 1500 1750
m/z
MS3
y10
942.48 m/z, 3+, 2,824.42 Da, (Parent Error: -6.5 ppm)
G L N S S M D M A R L P S PT+18 K E K P P P P P P G G G K
K G G G P P P P P P K E K T+18 P S P L R A M D M S S N L G
Relative Intensity
100%
75%
50%
25%
0%
y22+2H
y23+2H
y10+2H
b7+1
y4 b5 y6 y7 b7 y9 y10 b12 b13 b14 b17 b18
0 1000 2000
m/z

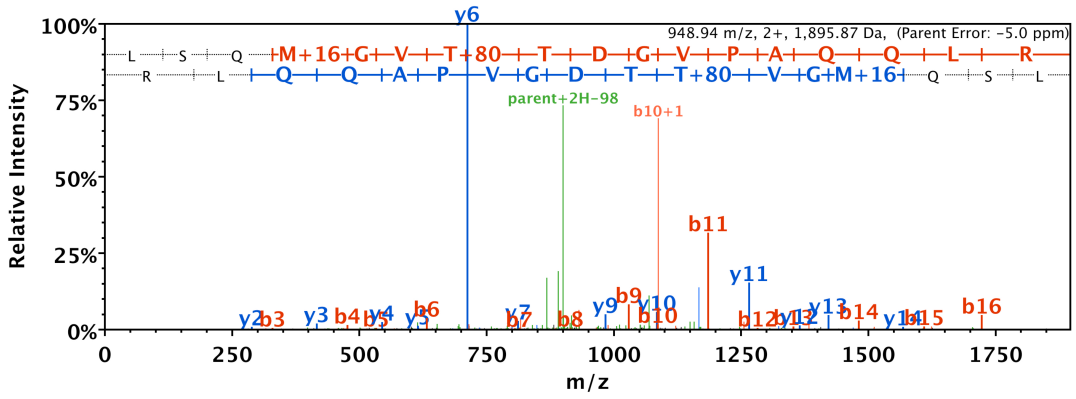
MS2 **S802**


parent+2H-98
1,033.92 m/z, 2+, 2,065.82 Da, (Parent Error: -1.2 ppm)
S S+80 P A Y C+57 T S S S D I T E P E Q K
K Q E P E T I D S S S T C+57 Y A P S+80 S
Relative Intensity
100%
75%
50%
25%
0%
0
500
1000
1500
2000
m/z
y4
y13
b14
y6
y16+2H
y16+2H+1
y11
b14-H2O
b14-H2O
y12
y14
y5
b5
y7
y8
y10
b11
b12
b13
b16
b17
y3
b4
b6
b7
b8
y9
b10

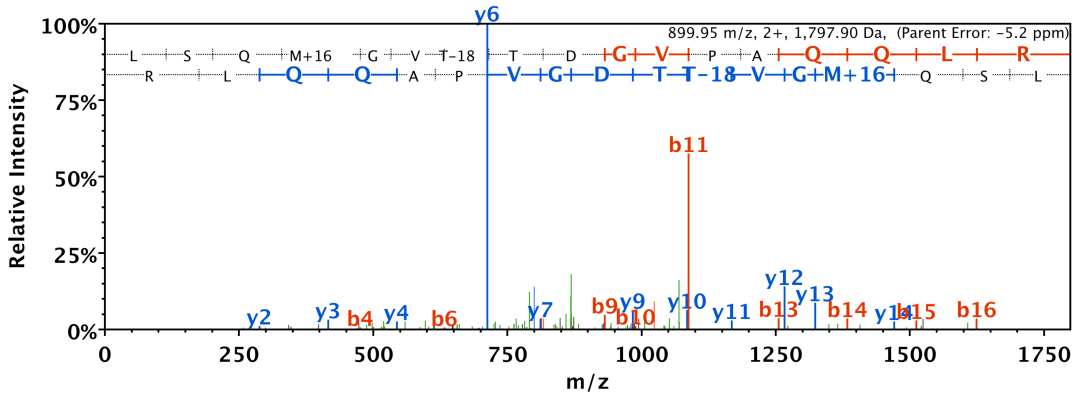
MS3


y11
984.93 m/z, 2+, 1,967.84 Da, (Parent Error: -1.2 ppm)
S Si-18 P A Y C+57 T S S S D I T E P E Q K
K Q E P E T I D S S T C+57 Y A P Si-18 S
Relative Intensity
100%
75%
50%
25%
0%
0
500
1000
1500
m/z
b14-H2O
y4
y6
b18-NH3-H2O+2H
y12
y14
y5
y7
y9
y10
b11
y13
b14
b16
y15
y2
b4
y3

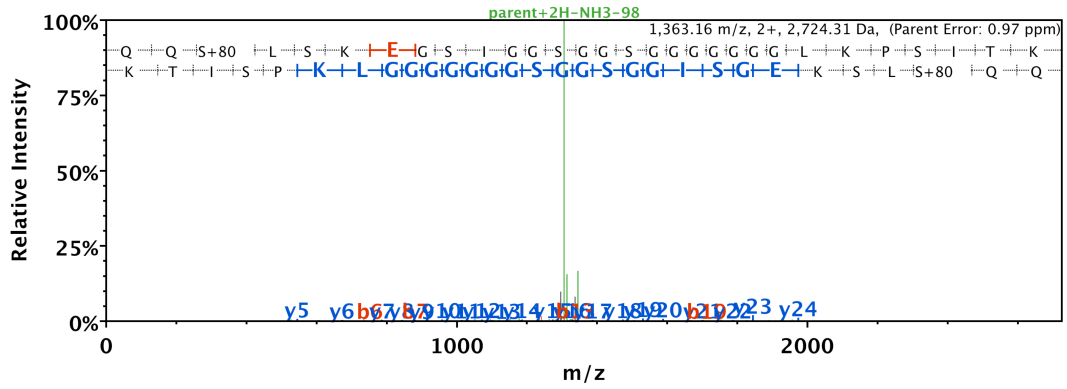
MS2 **T897**


948.94 m/z, 2+, 1,895.87 Da, (Parent Error: -5.0 ppm)
L S Q M+16 G V T+80 T D G V P A Q Q L R
R L Q Q A P V G D T T+80 V G M+16 Q S L
parent+2H-98
b10+1
Relative Intensity
m/z

MS3


899.95 m/z, 2+, 1,797.90 Da, (Parent Error: -5.2 ppm)
L S Q M+16 G V T-18 T D G V P A Q Q L R
R L Q Q A P V G D T T-18 V G M+16 Q S L
Relative Intensity
m/z

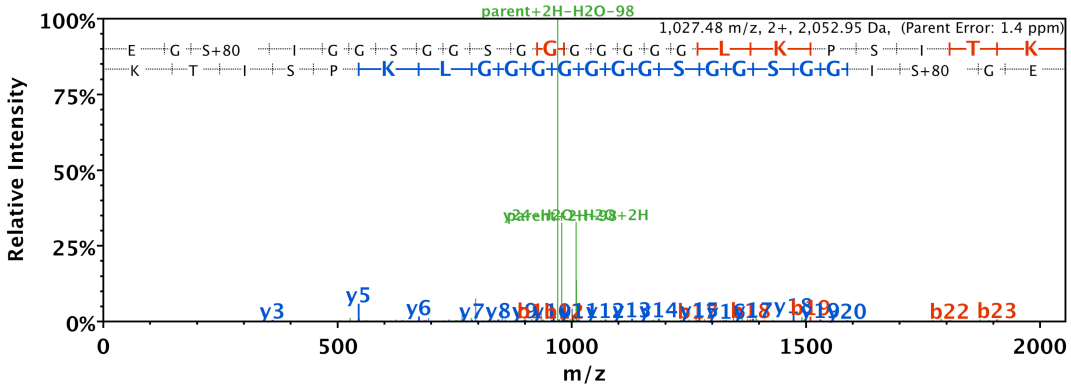
MS2 **S1093**



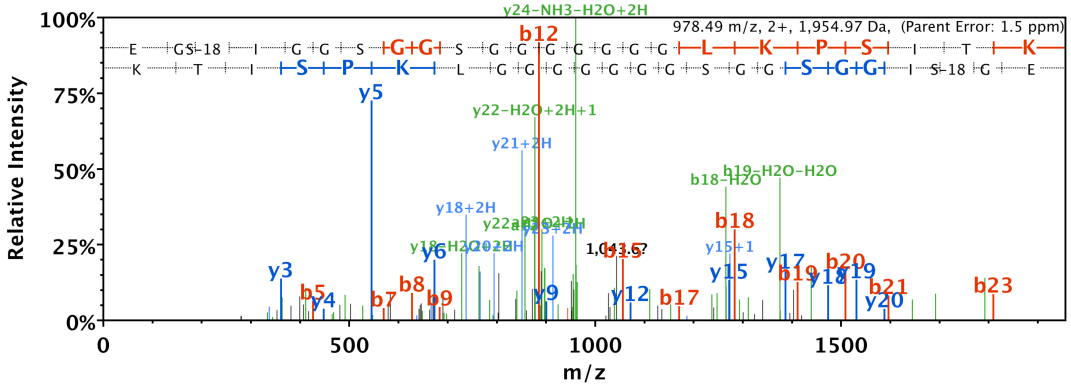
MS3

No MS3 Spectra

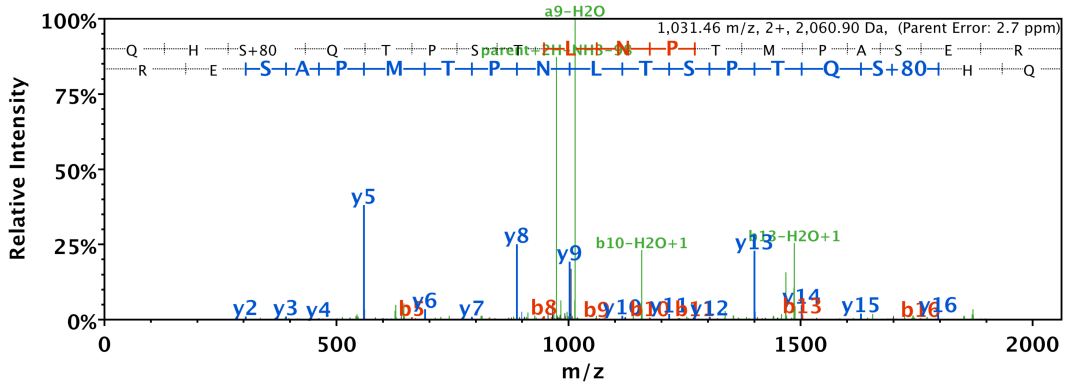
MS2 **S1099**



MS3



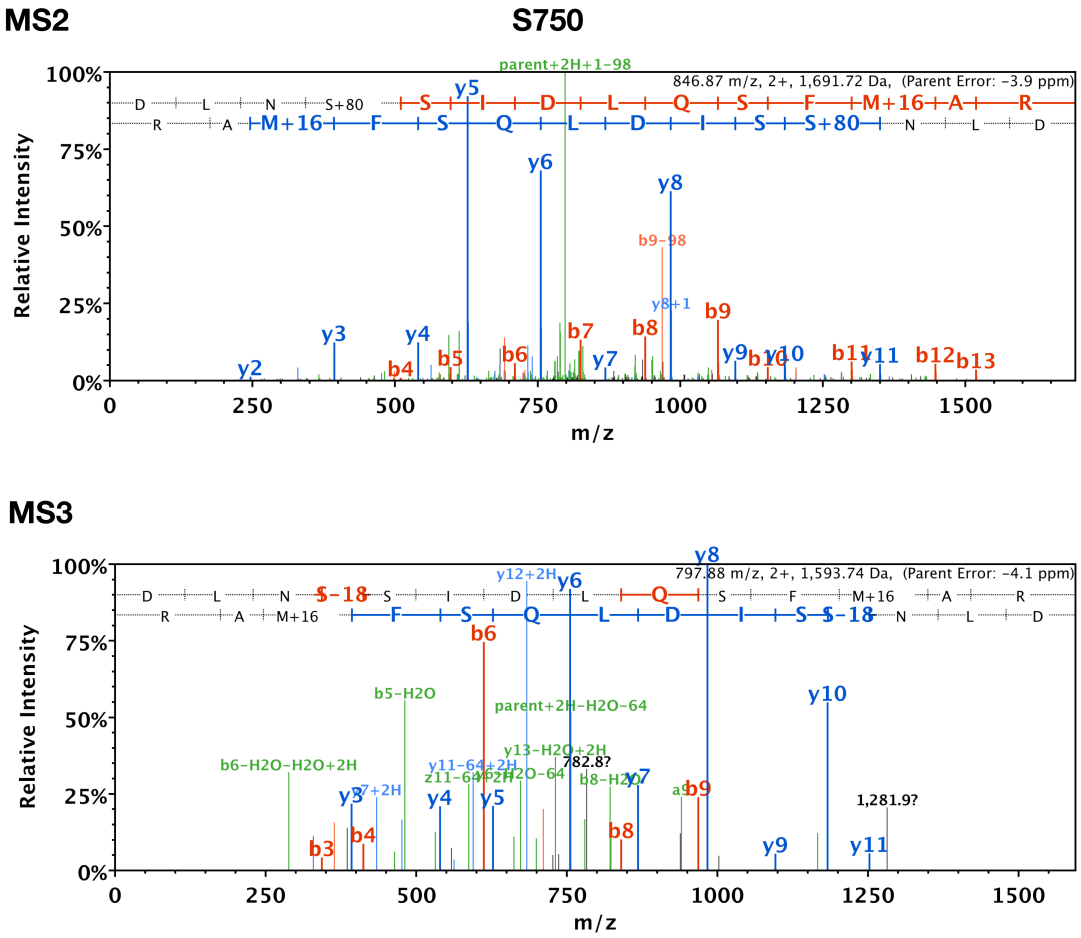
MS2 **S1123**

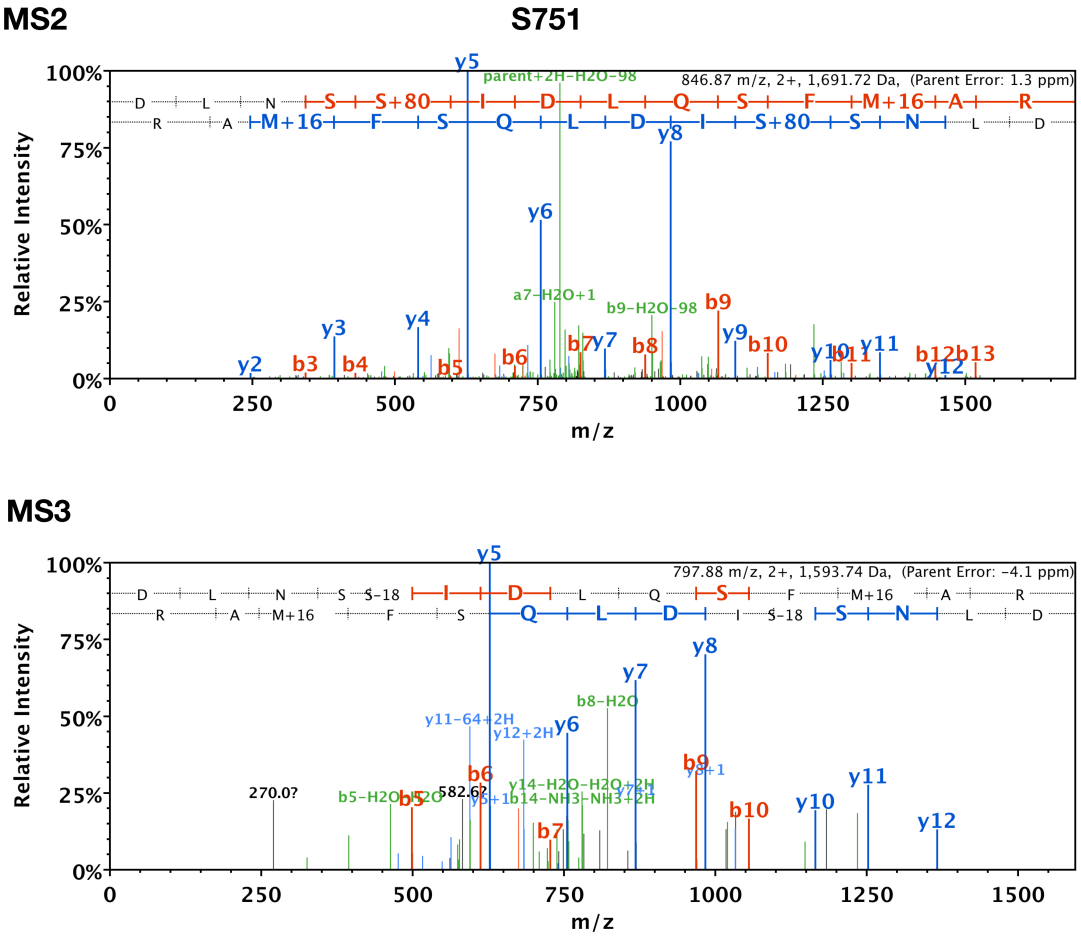


MS3

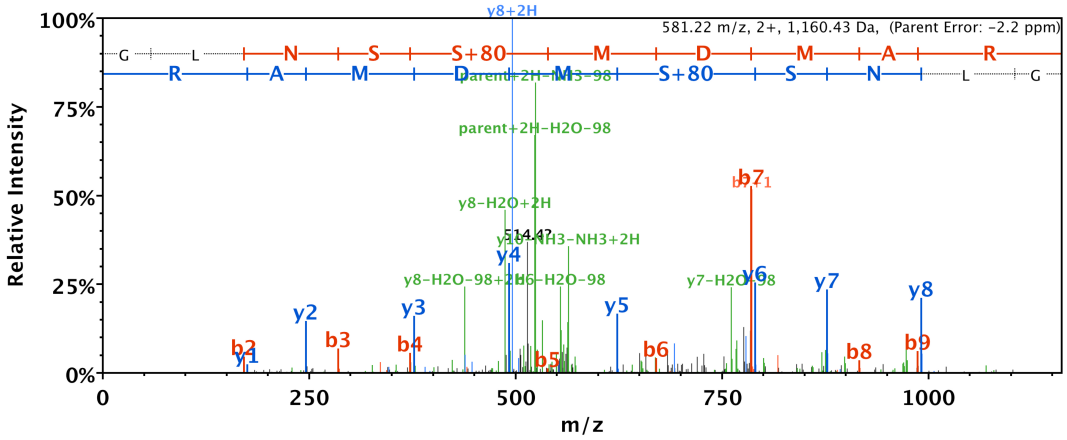
No MS3 Spectra

MASS SPECTRA FOR PLK2 PHOSPHORYLATION SITES IN SYNGAP




MS2
S751
846.87 m/z, 2+, 1,691.72 Da, (Parent Error: 1.3 ppm)
parent+2H-H2O-98
Relative Intensity
m/z
MS3
797.88 m/z, 2+, 1,593.74 Da, (Parent Error: -4.1 ppm)
270.0?
582.62

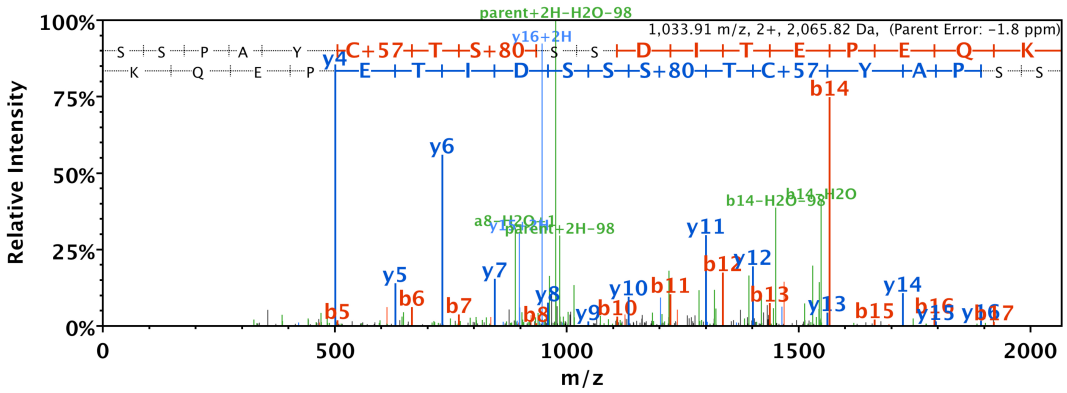
MS2 **S765**



MS3

No MS3 Spectra

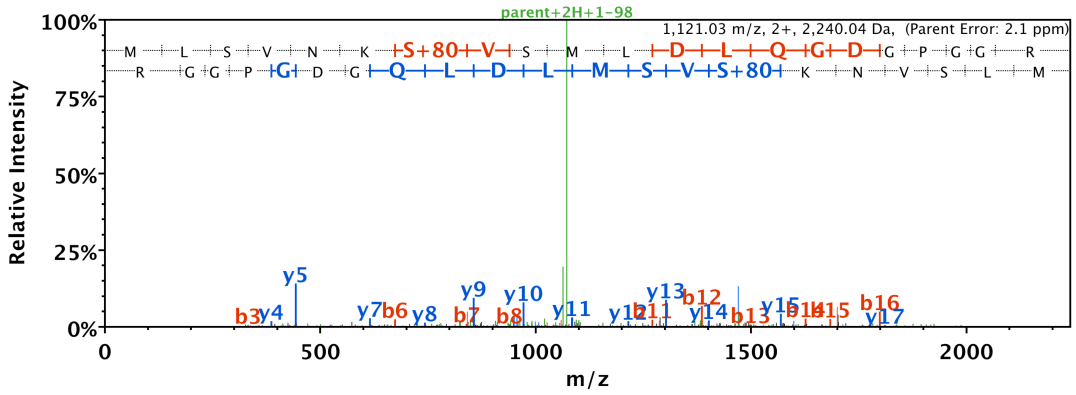
MS2 **T808**



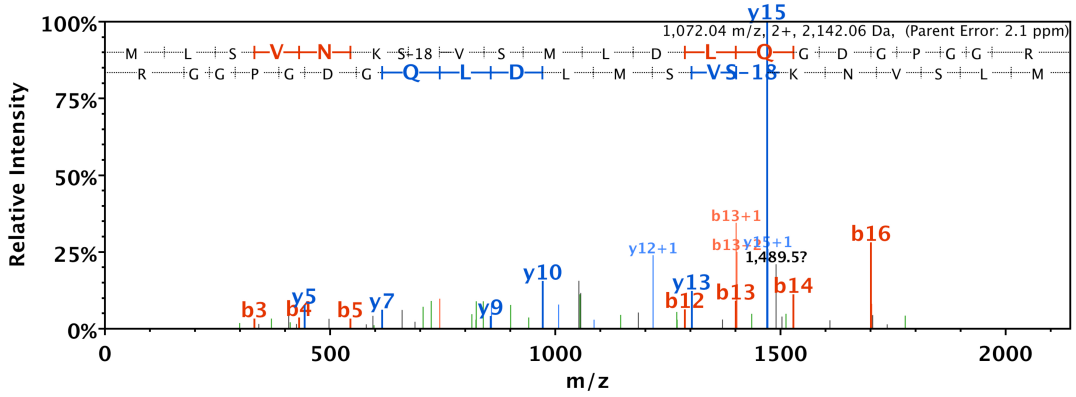
MS3

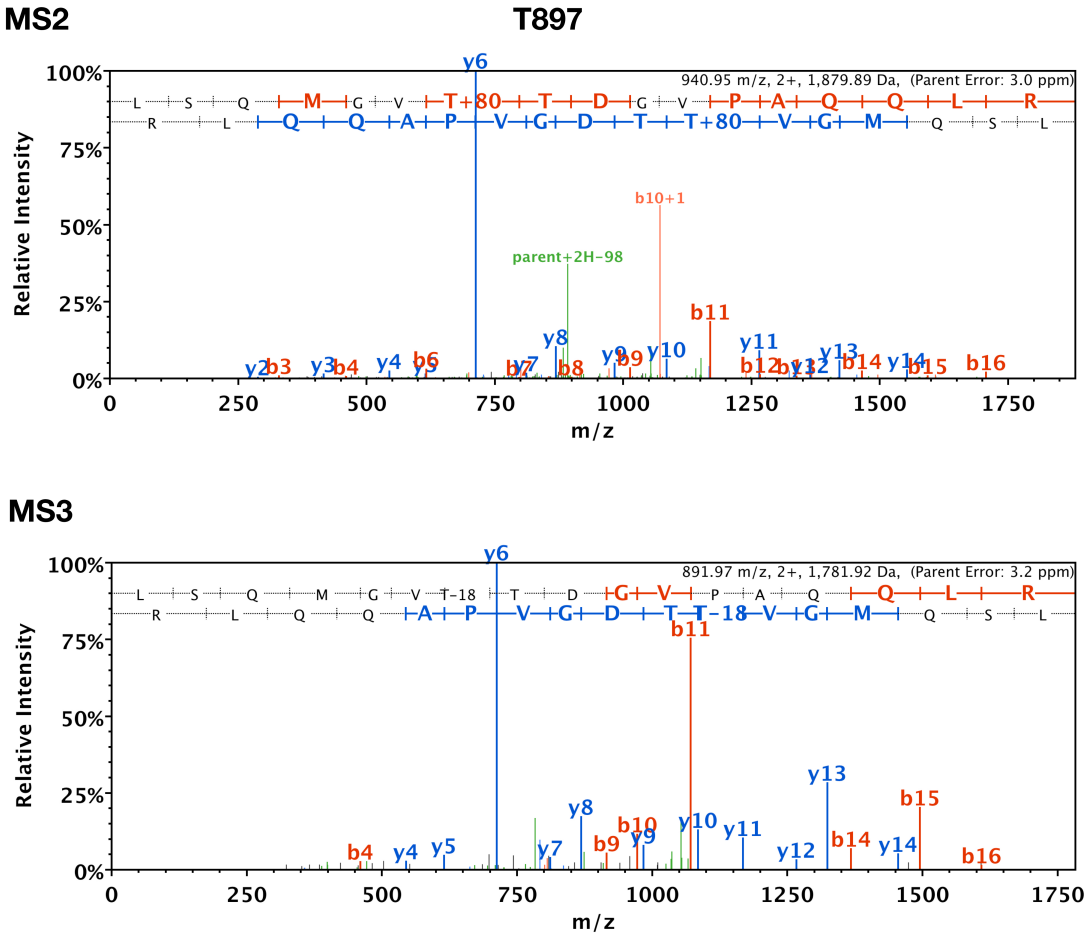
No MS3 Spectra

MS2 **T825**

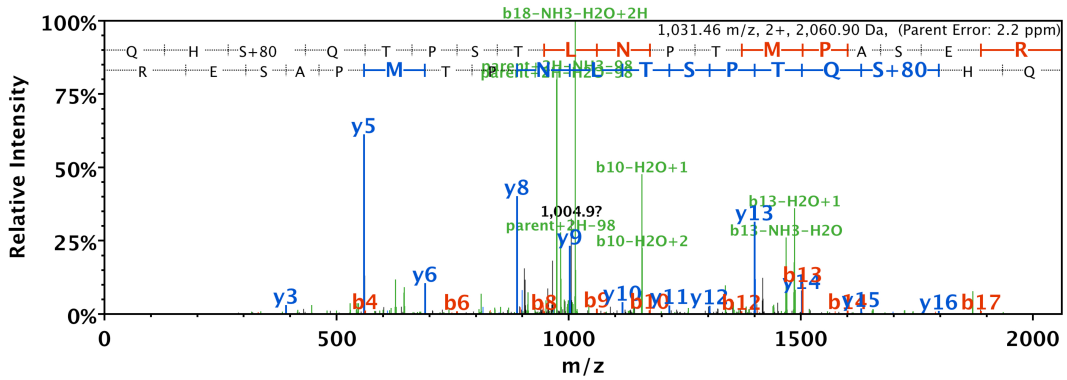

parent+2H+1-98
1,121.03 m/z, 2+, 2,240.04 Da, (Parent Error: 2.1 ppm)
M L S V N K S+80 V S M L D L Q G D G P G G R
R G G P G D G Q L D L M S V S+80 K N V S L M
Relative Intensity
100%
75%
50%
25%
0%
0
500
1000
1500
2000
m/z
y5
b3
y4
y7
b6
y8
b7
y9
b8
y10
y11
y13
b12
y14
b13
y15
b16
y17

MS3


y15
1,072.04 m/z, 2+, 2,142.06 Da, (Parent Error: 2.1 ppm)
M L S V N K S-18 V S M L D L Q G D G P G G R
R G G P G D G Q L D L M S V S-18 K N V S L M
Relative Intensity
100%
75%
50%
25%
0%
0
500
1000
1500
2000
m/z
b13+1
y12+1
b13y15+1
1,489.5?
b16
y10
y13
b12
b13
b14
y5
b3
b4
b5
y7
y9


MS2
T897
940.95 m/z, 2+, 1,879.89 Da, (Parent Error: 3.0 ppm)
L S Q M G V T+80 T D G V P A Q Q L R
R L Q Q A P V G D T T+80 V G M Q S L
y6
b10+1
parent+2H-98
b11
y8
y11
y13
y10
y9
b14
y14
b15
b16
y2
b3
y3
b4
y4
b6
y5
b7
b8
b12
y12
Relative Intensity
100%
75%
50%
25%
0%
0
250
500
750
1000
1250
1500
1750
m/z
MS3
891.97 m/z, 2+, 1,781.92 Da, (Parent Error: 3.2 ppm)
L S Q M G V T-18 T D G V P A Q Q L R
R L Q Q A P V G D T T-18 V G M Q S L
y6
b11
y13
b15
y8
b10
y10
y11
b4
y4
y5
y7
b9
y9
y12
b14
y14
b16
Relative Intensity
100%
75%
50%
25%
0%
0
250
500
750
1000
1250
1500
1750
m/z

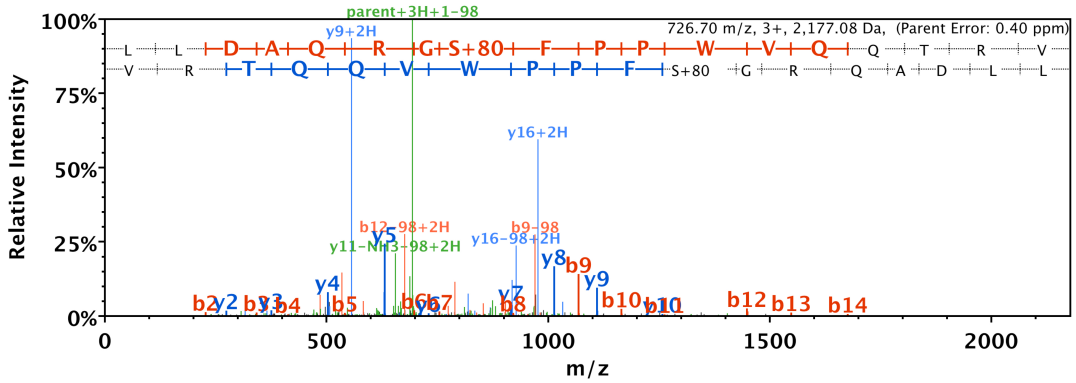
MS2 **S1123**



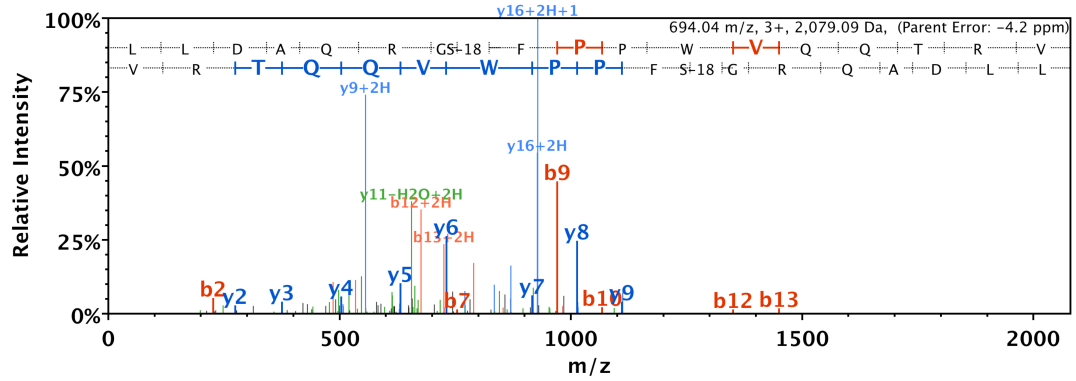
MS3

No MS3 Spectra

MS2 **S1283**



MS3



A p p e n d i x I I

PUBLICATIONS

AUTHOR'S COMMENTS

This section contains all of the published papers I have co-authored during my tenure at Caltech.

Appendix II:
Publications

Journal of Biomolecular Techniques
16:248–255 © 2005 ABRF

Automated Template Quantification for DNA Sequencing Facilities

Kathryn M. Ivanetich,[1] Wilson Yan,[1] Kathleen M. Wunderlich,[1] Jennifer Weston,[1] Ward G. Walkup IV,[2] and Christian Simeon[1]

[1]Biomolecular Resource Center and Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, California, [2]Graduate Department of Biochemistry and Molecular Biophysics, California Institute of Technology, Pasadena, California

The quantification of plasmid DNA by the PicoGreen dye binding assay has been automated, and the effect of quantification of user-submitted templates on DNA sequence quality in a core laboratory has been assessed. The protocol pipets, mixes and reads standards, blanks and up to 88 unknowns, generates a standard curve, and calculates template concentrations. For pUC19 replicates at five concentrations, coefficients of variance were 0.1, and percent errors were from 1% to 7% (n = 198). Standard curves with pUC19 DNA were nonlinear over the 1 to 1733 ng/µL concentration range required to assay the majority (98.7%) of user-submitted templates. Over 35,000 templates have been quantified using the protocol. For 1350 user-submitted plasmids, 87% deviated by ≥ 20% from the requested concentration (500 ng/µL). Based on data from 418 sequencing reactions, quantification of user-submitted templates was shown to significantly improve DNA sequence quality. The protocol is applicable to all types of double-stranded DNA, is unaffected by primer (1 pmol/µL), and is user modifiable. The protocol takes 30 min, saves 1 h of technical time, and costs approximately $0.20 per unknown.

KEY WORDS: DNA, plasmid, quantification, automation, PicoGreen, core facility.

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO: Kathryn Ivanetich, Biomolecular Resource Center, University of California San Francisco, Surge 104, 90 Medical Center Way, San Francisco, CA 94143–0541 (phone: 415-476-1047, 415-514-0102; fax: 415-476-7974; email: kathyi@cgl.ucsf.edu).

Templates submitted to biotechnology core facilities are typically of diverse vector types, and have been purified and quantified by diverse methods. Although this facility requests templates at 500 ng/µL, in preliminary experiments, template concentrations ranged from 0 to 6000 ng/µL (J. O'Shaughnessy, Hoechst 33258 assay, unpublished data). There is anecdotal evidence that template concentration can affect automated DNA sequencing success rates[1,2] and that high template concentrations can impact capillary viability and contaminate adjacent lanes on slab gel sequencers.

A number of methods are available to quantify double-stranded (ds) DNA. The absorbance at $\lambda =$ 260 nm is relatively insensitive (an A_{260nm} of 0.1 corresponds to 5 µg/mL dsDNA)[3] and nonspecific since numerous organic molecules and biopolymers including phenol, nucleotides, single-stranded (ss) DNA, protein, and RNA absorb at 260 nm. Fluorimetric dye binding assays for dsDNA have increased sensitivity and specificity. The ethidium bromide assay has higher sensitivity for dsDNA (1 µg/mL) than A_{260nm}, but the fluorescence yield is constant over a limited dye/DNA ratio, and the reagent fluoresces with RNA and ssDNA.[4–6] The Hoechst 33258 assay exhibits improved sensitivity (ca. 10 ng/mL) and specificity for dsDNA, but assays at two salt concentrations are required for high specificity.[7,8]

Compared with the aforementioned methods, the PicoGreen dye binding assay for dsDNA has greater sensitivity, specificity, and linear range.[9,10] The PicoGreen assay is sensitive to 25 pg dsDNA/mL, which is 400-fold more sensitive than Hoechst 33258 dye. The PicoGreen assay has high specificity for dsDNA versus RNA or ssDNA, and is unaffected by protein, phenol, and numerous other reagents. For calf thymus DNA, the PicoGreen assay has a linear range of four orders of magnitude with a single dye concentration.[9]

There have been few reports on automation of DNA quantification with application to large numbers of user-submitted samples in a core or high throughput facility. Haque et al. compared PicoGreen, A_{260nm}, and quantitative PCR (for a specific gene target) for quantification of a narrow range of dsDNA concentrations (up to 20 ng/μL) for a high-throughput genomics setting; they concluded that the PicoGreen assay had advantages of specificity and sensitivity.[11] Tecan Robotics and Beckman Coulter offer automated DNA quantification protocols for integrated absorbance or fluorimetry plate readers.[12–15] However, there are no published reports of application of these protocols to large numbers of samples.

An automated method to quantify plasmid DNA based on the PicoGreen assay has been developed, and over 35,000 DNA templates have been quantified. The protocol provides accurate determination of a broad range of DNA template concentrations, i.e., 30 to 1733 ng/μL (initial concentration). Based on 418 DNA sequencing runs, the quantification of user-submitted templates by the reported protocol significantly improved DNA sequence quality.

MATERIALS AND METHODS

Materials

The PicoGreen dsDNA Quantitation Kit [PicoGreen (320 μM in DMSO), 20× TE buffer (200 mM Tris-HCl/20 mM EDTA buffer, pH 7.5), and lambda DNA (0.1 μg/μL TE)] was from Molecular Probes (Eugene, OR). pUC19 (1 μg/μL) was from New England Biolabs (Beverly, MA). Corning-Costar clear v-bottom and black flat bottom 96-well plates were purchased from PGC Scientific (Frederick, MD) or Greiner (Pleasanton, CA). Pipet tips were from Robbins Scientific (Sunnyvale, CA) or Molecular BioProducts (San Diego, CA). Liquinox detergent was from Alconox, Inc. (New Hyde Park, NY).

Instrumentation

The Beckman Coulter (Palo Alto, CA) Biomek 2000 robotic workstation included left- and right-side modules, and the pipette and gripper tools and holders shown in Figure 1. The BMG FLUOStar 97 fluorescence plate reader (Durham, NC) was attached to the left side of the Biomek deck. The FLUOStar 97 was operated at a fixed gain of 20 with BMG 485-12 excitation and 520–35 emission filters. An IBM 300GL computer with Microsoft Windows NT operating system, Microsoft Excel 97, Bioworks v.3.1 and SILAS software (Beckman Coulter), and FLUOStar Galaxy software v.4.30-0 (BMG), control the Biomek and FLUOStar 97.

Protocol Setup

The setup of the deck for the Biomek protocol is shown in Figure 1. The following solutions are loaded onto the Biomek: pUC19 (100 μL, 0.1 μg/μL TE) in a 1.5-mL microfuge tube and an empty 1.5-mL microfuge tube are placed in positions A1 and B1, respectively, of the microfuge 24 holder. For some experiments, lambda DNA (0.1 μg/μL TE) replaced pUC19 as DNA standard. TE and PicoGreen (1.6 μM in TE) are added to the quarter reservoirs. Approximately 10 μL of the templates to be quantified are placed in columns 2 through 12 of the PE 9600 plate. The read plate wells contain, in a total volume of 200 μL, 0.8 μM PicoGreen, TE, and up to 0.43 ng/μL (87 ng total) pUC19 DNA. The wells containing blanks, unknowns, and standards, and the concentrations of the latter are entered into the FLUOStar Galaxy software. FloPatterns 1 and 2 of the Biomek Bioworks software are modified for the number of templates to be quantified.

Plasmid Quant Protocol

The Plasmid Quant protocol generates tables of fluorescence intensities (uncorrected and corrected for blanks; not shown), a standard curve (Fig. 2), and a table of the calculated concentrations of unknowns. DNA concentrations are given for the original solutions. The pUC standards correspond to 1733, 1200, 700, 400, and 100 ng/μL of DNA, and a 1-ng/μL point is included for curve-fitting purposes. DNA unknowns of higher concentrations were flagged by the protocol, manually diluted five-fold, and reassayed.

An overview of the protocol is given in Table 1. The protocol is in 96-well format and performs all pipetting, mixing, plate reading, and data handling. The reproducibility and accuracy of the Biomek MP20 tool and the manual Rainin Pipet-Lite L8–10 multichannel pipet (Oakland, CA) in pipeting 1-μL aliquots of pUC19 into 99 μL of TE were compared. For the MP20 robotic pipeting, the coefficient of variance (CV) ranged from 30% to 42%, and percent error from theoretical was from 5% to 32% (mean 13.6%). The manual multichannel pipettor (CV ca. 10%) was found to be three-fold more reproducible than the Biomek robot for pipeting 1 μL, if the remainder of the protocol was pipetted robotically. In all further experiments, 1-μL volumes were pipetted manually with the multichannel pipettor.

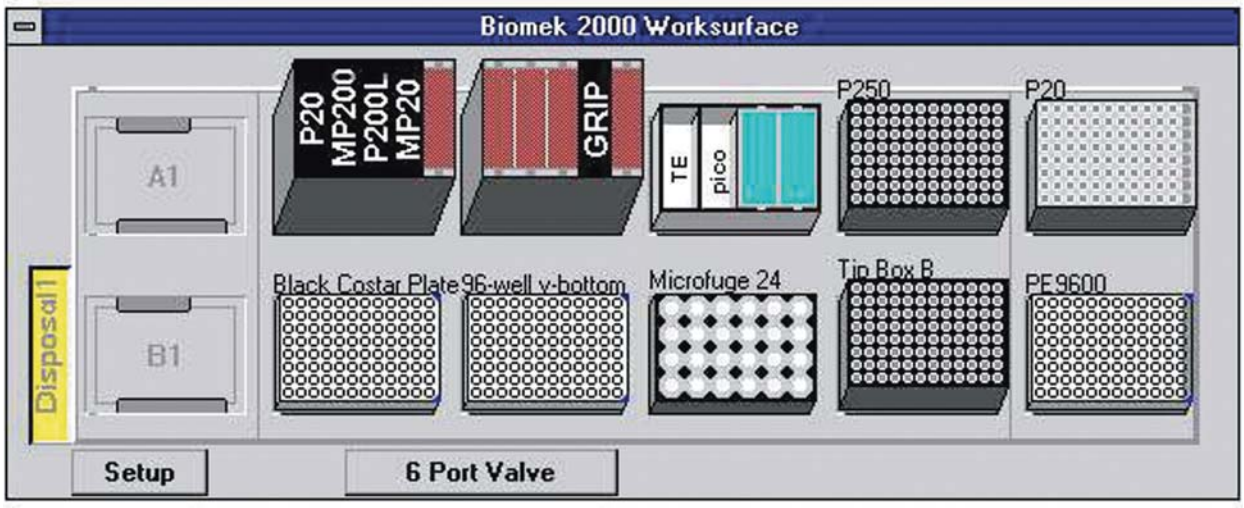


FIGURE 1

Set up of the deck of the Biomek 2000.

The protocol can be modified by the user at numerous levels. The number of columns to be pipetted can be modified based on the number of templates to be assayed. The deck setup and pipeting parameters can be modified, as can be the range of the standard curve and the dilution of standards and unknowns. In addition, the protocol can be truncated at the end of the pipeting and mixing to produce read plates suitable for a nonintegrated fluorimetry plate reader.

DNA Sequencing

Of the 35,000 templates quantified with the protocol, approximately 18,000 were sequenced with BigDye chemistry version 2.0, 4000 with BigDye v.3.0, and 13,500 with BigDye v.3.1. Data are presented only for DNA sequenced with BigDye Terminator chemistry v.3.1 on an Applied Biosystems (ABI) PRISM 3700 capillary sequencer with POP6 polymer (ABI, Foster City, CA), as described previously.[16] Reaction mixtures contained 185–500 ng template DNA (unless otherwise indicated), primer (3.5 pmol), BigDye Terminator Master Mix (1.2 μL), 5× Sequencing buffer (1.8 μL), and water in a total volume of 10 μL. One M13F primer/pGEM sequencing reaction was run per 96 well plate. Average and 1st Phred <20 scores were generated automatically by dnaLIMS software (dnaTools Inc., Ft. Collins, CO). The Average Phred score is calculated over the entire sequencing run and the 1st Phred <20 score is the first base that the Phred score drops below 20. Unsuccessful runs were defined by Average Phred scores below 25 or 1st Phred <20 scores below 300.

Statistical analysis was performed using the Student's *t*-test for unpaired data, unless otherwise indicated. A significant difference between means was taken as $p < 0.05$.

RESULTS AND DISCUSSION

Assay Validation

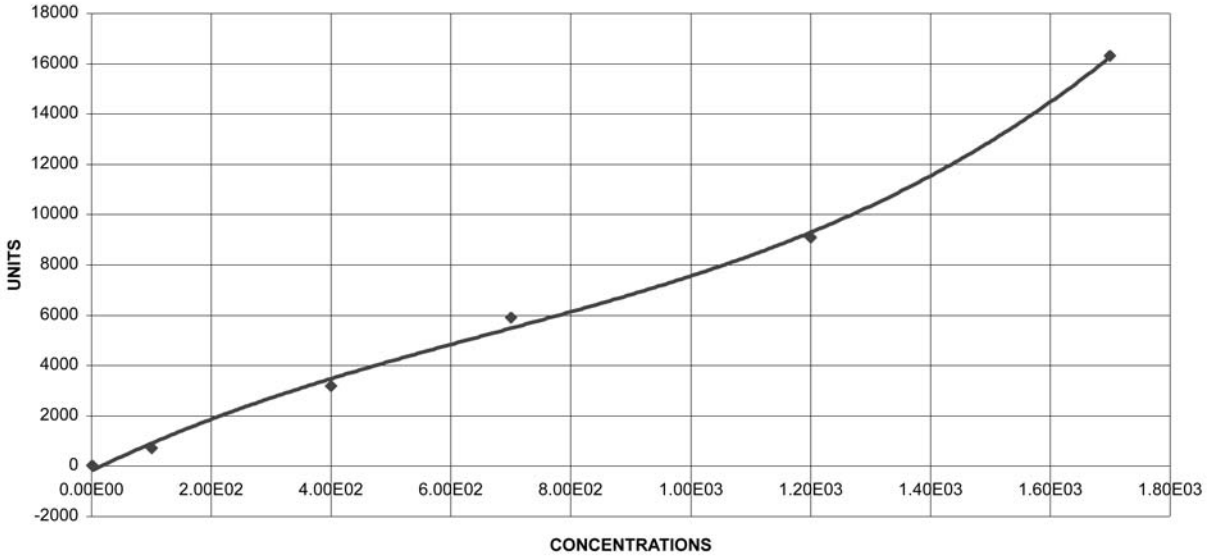
The choice of the DNA standard for the PicoGreen assay was not trivial since different types of dsDNA had variable fluorescent yields with PicoGreen. In addition, although standard curves for several types of DNA including bacteriophage lambda and calf thymus dsDNA were linear, those for pUC19 and øX174 were nonlinear.[9] Since most templates quantified in this facility are plasmids, pUC19 was chosen as the DNA standard. Over the concentration range (up to 1733 ng/μL) needed to assay a high percentage of user-submitted templates (Fig. 3), the standard curve for pUC19 remained nonlinear in over 25 interim protocols, with varying concentrations of standards and PicoGreen (data not shown) and in the final protocol (Fig. 2).

Replicate determinations (n = 98) at two concentrations of pUC19 DNA (294 and 571 ng/μL) had CVs and % errors of approximately 10% (Table 2). Quantification of replicates of two concentrations of pUC19 (n = 32) using a lambda DNA standard curve resulted in CVs (ca. 10%) similar to those found with a pUC19 standard curve, but % errors of approximately 30%, i.e., three-fold higher for the lambda DNA standard curve. In order to minimize systematic error in the PicoGreen assay, it is essential to choose the most

FLUOstar Galaxy	Test Name:	FLO LAYOUT	2004/09/02	12:07:03
BMG Labtechnologies	ID 1,2,3:	090204		

Polynomial Regression: Y = b + c1x + c2x_ + c3x_ Evaluation Type: Sum

b	-1.83E2
c1	1.17E1
c2	-7.90E-3
c3	4.00E-6
r	0.99895



3rd Polynomial

Standard Concentrations

	1	2	3	4	5	6	7	8	9	10	11	12
A	1700											
B	1200											
C	700											
D	400											
E	100											
F	1											
G												
H												

Calculated Concentrations

	1	2	3	4	5	6	7	8	9	10	11	12
A	1688.8635	176.90208	366.85406	258.84966	667.76021	484.82597	91.14135					
B	1169.018	124.07757	376.33732	556.56602	559.28458	636.84455	105.95516					
C	763.35318	1461.9314	668.06784	351.2768	637.45941	645.91631						
D	360.35761	598.81532	645.76251	1216.0619	401.56932	119.05628						
E	82.217665	353.50678	542.41478	380.36912	850.90018	160.26861						
F	18.33003	111.88256	155.84917	367.38613	379.69613	128.32077						
G		337.8683	204.07283	301.43749	517.92938	146.04683						
H		415.22455	298.20469	613.05004	299.19821	187.96733						

FIGURE 2

Typical output: pUC19 standard curve and calculated concentrations of standards and unknowns.

TABLE 1

Overview of the Plasmid Quant Program Protocol

1. Dilute the pUC19 DNA standard (100 ng/mL) to 20 ng/mL (deck position B4)(P20 and P200 tools).
2. Pipet blanks and dilutions of the pUC19 standard (<65 mL pUC19) in the wells of column one of the 96 well dilution plate (deck position B3).
3. Pipette 99 mL of TE (MP200 tool) into the columns of the 96-well dilution plate designated to contain unknowns (deck position B3).
4. Pipette 95 mL of TE (MP200 tool) into the columns of the 96-well read plate designated to contain blanks, standards, and unknowns (deck position B2).
5. Transfer 1 mL of unknowns (MP20 tool) from the PE9600 plate (deck position B6) to the dilution plate. (Alternatively, pause robot and manually transfer the unknowns with an 8-channel pipettor, and restart the robot.)
6. Pipet 100 mL of PicoGreen (MP200 tool) into the designated columns of the read plate.
7. Pipet 5 mL of diluted standards, blanks, and unknowns (MP20 tool) from the dilution plate into the corresponding columns of the read plate.
8. Open the FLUOStar drawer, move the read plate into the FLUOStar with the Gripper tool, wait 5 min, and read the plate.
9. Perform all data handling: correct fluorescence intensities for blanks, generate the standard curve, determine the concentrations of all unknowns within the range of the standard curve, and flag outliers.
10. Remove plate from reader.

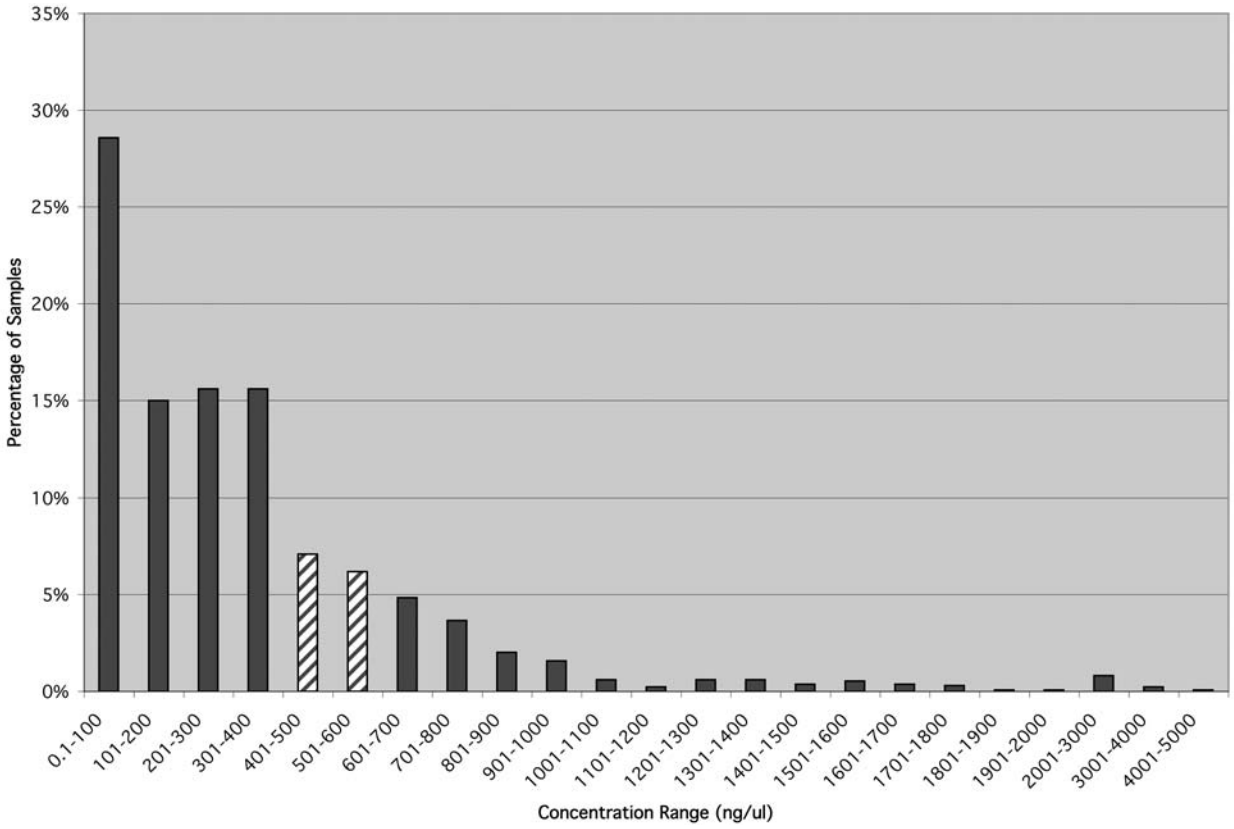


FIGURE 3

Distribution of data for plasmid templates reportedly submitted to the biotechnology facility at 500 ng/µL. Cross-hatched bars denote template concentrations within 20% of the requested concentration.

Table 2

Replicate Assays on pUC19 in the Presence and Absence of M13 Primer and on Lambda DNA

	pUC19		pUC19 – Primer	pUC19 + Primer[a]	Lambda DNA	
Theoretical Conc. (ng/µL)	294	571	147	147	480	917
Experimental Conc. (ng/µL) Mean	314	517	156	160	529	849
SD (ng/µL)	31	58	17	9	62	93
Percent Error (%)	7.9	8.7	6.1	8.7	11	6.6
C.V. (%)	10	11	11	5	12	11
Minimum (ng/µL)	223	399	130	138	402	684
Maximum (ng/µL)	390	727	190	179	694	1,051
Count	98	98	22	24	47	32

[a]1 pmol of M13 (–20) universal primer per 50 ng pUC19.

appropriate DNA standard for the type of DNA assayed, namely pUC19 for plasmid quantification.

Data were compiled for replicates for the calculated concentrations and fluorescence intensities for each of the five concentrations of pUC19 used to generate the standard curve. The theoretical concentrations and the experimentally determined mean, standard deviation (SD), minimum, maximum, CV, % error, and n values are given in Table 3. The calculated concentrations of the standards agreed closely with the theoretical values for concentrations from 100 ng/µL to 1733 ng/µL (% error of <1% to 7%). The CVs increased from 2% to 15% as the concentration of the standard decreased from 1733 to 100 ng/µL. In contrast, the fluorescence intensity for each pUC19 concentration varied strikingly from run to run. For the 1733 ng/µL standard, the fluorescence intensity had a mean of 35,400, but ranged from 18,000 to 64,500 (Table 3). Similar broad ranges of fluorescence intensity were observed at each concentration of pUC19, with CVs ranging from 34% to 47%. The CVs for fluorescence intensity were 2.5-fold to 20-fold greater than the CVs for pUC19 concentration, indicating that a standard curve should be generated for each run.

The assay was not developed for accuracy for DNA concentrations below 30 ng/µL since the facility has anecdotal evidence from dRhodamine and older Dye Terminator chemistries run on slab gel and capillary sequencers that DNA sequencing reactions on templates submitted below 60 ng/µL have a high probability of failure (data not shown).

Template Quantification

Over 35,000 templates have been quantified with the Plasmid Quant protocol. Representative data for 1350 plasmids are given in Figure 3. All templates were ostensibly quantified by users and submitted at 500 ng/µL. Only 13% of the templates were found to have concentrations within 20% of the target concentration, i.e., between 401 and 600 ng/µL (cross-

Table 3

Data from pUC19 Standards Used for Preparation of the Standard Curve

(ng/µL	Mean	SD	CV	Min	Max	n	% Error	Mean	SD	CV	Min	Max	n
1733	1732	29	2%	1664	1794	37	0.1%	35413	12089	34%	17958	64458	24
1200	1187	47	4%	1057	1282	46	1.1%	22647	10731	47%	11530	54860	23
700	704	35	5%	644	809	47	–0.5%	12289	5778	47%	7293	34189	24
400	403	34	8%	335	487	47	–0.7%	6679	2564	38%	3345	14988	24
100	106	16	15%	81	144	47	–6.4%	1711	642	38%	529	3687	24
1	8	7	85%	0	24	40	–684%	335	154	46%	120	687	23

hatched). Approximately 28% of the submitted templates were below 100 ng/μL, i.e. more than five-fold below the target concentration. Fifteen percent of the submitted templates were in each of the following ranges: 101–200, 201–300, and 301–400 ng/μL, and 2% to 5% of the submitted templates were in the 100 ng/μL concentration ranges between 600 and 1000 ng/μL. Above 1000 ng/μL, each 100 ng/μL range accounted for less than approximately 1% of the total.

With a standard curve up to 1733 ng/μL, 98.7% of the 1350 templates assayed fell within the standard curve, and only 1.3% or 17 templates required dilution and re-assay. If the concentration range of the standard curve were decreased, a greater percentage of samples would require dilution and re-reading. For example, with a standard curve up to 500 ng/μL, 24%, or 318 templates, would require dilution and re-assay. M13 Universal primer (1 pmol/μL), which is at a three-fold higher concentration than that used in sequencing reactions, did not affect plasmid quantification (Table 2).

Effect of Template Quantification on DNA Sequencing

The effect of template quantification on the quality of DNA sequencing data for user-submitted templates with inserts was assessed in three experiments. In the first experiment, 99 templates were sequenced without quantification. Seventy four percent of the templates had Average Phred scores over 25 (mean scores of 44 ± 9) and were classified as successful, while 26% had Average Phred scores below 25 (mean scores of 12 ± 5) and were considered unsuccessful. In contrast, 81% of 183 randomly selected templates quantified by the protocol with the template concentration in the sequencing reaction mixture optimized, had Average Phred scores greater than 25 (mean scores of 49 ± 7), i.e., a 10% increase in DNA sequencing success rate for quantified templates.

Two subsequent experiments focused on templates that quantified below 60 ng/μL or over 800 ng/μL. Each template was sequenced at two concentrations, with all other reaction conditions held constant. This generated results equivalent to (i) no in-house template quantification and (ii) adjusting template concentration for quantification as much as possible.

For 40 templates that quantified below 60 ng/μL, each template was run (i) as if it was at the requested concentration of 500 ng/μL (0.5 μL added to reaction mixtures), and (ii) optimizing the amount of DNA in the reaction mixture based on quantification (6 μL added to reaction mixtures). Option (i) resulted in 15 ± 11 ng DNA in reaction mixtures, and only 12.5% or 5 of 40 successful sequencing runs (Average Phred scores of 11 ± 9). Option (ii) resulted in 173 ± 132 ng DNA in reaction mixtures, and an 87.5% success rate or 35/40 successful runs (Average Phred scores of 36 ± 11). Quantification of this set of templates increased Average Phred scores four-fold (4.4 ± 1.8) ($p < 0.001$). Analogous results were obtained with 1st Phred <20 scores or average signal intensities (data not shown).

For 28 templates that quantified between 814 and 1,952 ng/μL, each template was run (i) as if it was at the requested concentration of 500 ng/μL (0.5 μL added to reaction mixtures), and (ii) optimizing the amount of DNA added to reaction mixtures (1 μL of a 1:3 dilution). For options (i) and (ii), reaction mixtures contained 723 ± 145 ng versus 482 ± 97 ng DNA, respectively. Option (i) had a success rate of 35.7% (10/28 templates sequenced successfully) with Average Phred scores of 18 ± 12, while option (ii) had a success rate of 78.6% (22/28) with Average Phred scores of 38 ± 14. Quantification of high concentration templates (800 ng/μL) improved Average Phred scores three-fold (3.0 ± 2.1) ($p < 0.001$). Analogous results were obtained with 1st Phred <20 scores (data not shown).

Since 26% of the user-submitted templates quantified by the Plasmid Quant protocol were at concentrations below 60 ng/μL or above 800 ng/μL, with analysis of 35,000 samples, quantification would identify 9250 templates with a low probability of successful runs and save users and/or the facility the cost of these runs and troubleshooting.

Applicability to Other Types of dsDNA

The applicability of the protocol to other types of dsDNA was validated by substitution of lambda DNA (100 ng/μL) for the pUC19 standard in several runs. The resulting standard curves were linear ($r = 0.998$). For replicate assays at two concentrations of lambda DNA, CVs and % errors (ca. 10%) were comparable to those for pUC19 (Table 2).

CONCLUSIONS

Automation of template quantification saves approximately 1 h of technician's time per 88 templates quantified. The manual duties eliminated by the protocol include numerous pipetting and mixing steps, construction of the standard curve, and calculation of template concentrations. The savings in personnel time is equal to an annual savings of $4500, assuming

one full quantification run per day and salary and benefits of $35,000 annually. The automated protocol is twice as fast as the manual equivalent, requiring only 30 min per full run. The cost to run the protocol for the standard curve plus 88 samples is $18.74 for consumables (including $3.20 for the use of pUC19), which is equal to $0.21 per sample. The total cost can be reduced by washing the 96 well plates in Liquinox and reusing them.

In conclusion, the application of the Plasmid Quant protocol to user-submitted templates in a core facility setting improves the quality of DNA sequencing data and facilitates troubleshooting. The protocol has significant advantages over commercially available alternatives. Firstly, the choice of the appropriate type of DNA for standard, and the nonlinearity of standard curves with plasmid DNA has been extensively addressed in this protocol, but neither has been addressed in the commercially available protocols. Secondly, none of the commercially available protocols have standard curves beyond 500 ng/μL, which would necessitate dilution and re-assay of approximately 25% of submitted templates in a core facility setting. Finally, the Plasmid Quant protocol is available upon request at no charge, versus the cost of approximately $3000 for commercially available protocols.

ACKNOWLEDGMENTS

This work ws supported by NIH grants SIG S10RR15938 and S10RR13949 to KI for the Applied Biosystems 3700 DNA sequencer and Beckman Biomek 2000.

REFERENCES

1. Applied Biosystems. Automated DNA sequencing. Foster City, CA: Applied Biosystems, 2000.
2. Qiagen. The Qiagen guide to template purification and DNA sequencing. Chatsworth, CA: Qiagen, 1995.
3. Freifelder D. *Physical Biochemistry: Applications to Biochemistry & Molecular Biology*. San Francisco, CA: W.H. Freeman, 1982:494–536.
4. Dutton MD, Varhol RJ, Dixon DG. Technical considerations for the use of ethidium bromide in the quantitative analysis of nucleic acids. *Anal Biochem* 1995;230: 353–355.
5. Le Pecq JB, Paoletti CA. A new fluorimetric method for RNA and DNA determination. *Anal Biochem* 1966;17: 100–107.
6. Markovits J, Roques BP, Le Pecq JB. Ethidium dimer: A new reagent for the fluorimetric determination of nucleic acids. *Anal Biochem* 1979;94:259–264.
7. Weisblum B, Haenssler E. Fluorimetric properties of the bibenzimidzaole derivative Hoechst 33258, a fluorescent probe specific for AT concentration in chromosomal DNA. *Chromosoma (Berl.)* 1974;46:255–260.
8. Lebarca C, Paigen K. A simple, rapid, and sensitive DNA assay procedure. *Anal Biochem* 1980;102:344–352.
9. Singer VL, Jones LJ, Yue ST, Haugland RP. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. *Anal Biochem* 1997;249:228–238.
10. Ahn SJ, Costa J, Emanuel JR. PicoGreen quantitation of DNA: Effective evaluation of samples pre- or post-PCR. *Nucleic Acids Res* 1996;24:2623–2625.
11. Haque KA, Pfeiffer RM, Beerman MB, Struewing JP, Chanock SJ, Bergen AW. Performance of high-throughput DNA quantification methods. *BMC Biotechnol* 2003;3:20.
12. Tecan Company. Comparison of two different detection techniques for DNA quantification: Absorbance vs. fluorescence. *Application Note* 2000.
13. Jacotot L, Madarapu S, Ptylik M. Automating the normalization of DNA samples. *GEN Drug Discovery* 2000;20:32.
14. Kurapati R, Cook T, Calder M, Threadgill G. Automation of quantitation and normalization of DNA procedures on the Biomek 2000 Laboratory Automation Workstation. *Lab Automation.* Palm Springs, CA, 2001.
15. Kurapati R, Cook T, Calder M, Oda R, Threadgill G. Quantitation and normalization on the Biomek 2000 *Laboratory Automation Workstation* 2001;5(1):5.
16. O'Shaughnessy JB, Chan M, Clark K, Ivanetich KM. Primer design for automated DNA sequencing in a core facility. *Biotechniques* 2003;35:112–116, 118–121.

Available online at www.sciencedirect.com
ScienceDirect
Journal of Microbiological Methods 67 (2006) 507–526
Journal of Microbiological Methods
www.elsevier.com/locate/jmicmeth

Microbial source tracking by DNA sequence analysis of the *Escherichia coli* malate dehydrogenase gene

Kathryn M. Ivanetich [a,*], Pei-hsin Hsu [a], Kathleen M. Wunderlich [a], Evan Messenger [a], Ward G. Walkup IV [b], Troy M. Scott [c], Jerzy Lukasik [c], Jerry Davis [d,e]

[a] *University of California San Francisco, Biomolecular Resource Center and Department of Pharmaceutical Chemistry, Surge 104, 90 Medical Center Way, San Francisco, CA 94143-0541, USA*
[b] *Graduate Department of Biochemistry and Molecular Biophysics, California Institute of Technology, Pasadena, CA, USA*
[c] *Biological Consulting Services of North Florida, Inc., 4641 NW 6th Street, Suite A, Gainesville, FL 32609, USA*
[d] *Department of Geography and Human Environmental Studies, San Francisco State University, San Francisco, CA 94132, USA*
[e] *San Pedro Creek Watershed Coalition, Pacifica, CA 94404, USA*

Received 14 November 2005; received in revised form 25 April 2006; accepted 25 April 2006
Available online 14 September 2006

Abstract

Criteria for sub-typing of microbial organisms by DNA sequencing proposed by Olive and Bean were applied to several genes in *Escherichia coli* to identify targets for the development of microbial source tracking assays. Based on the aforementioned criteria, the *icd* (isocitrate dehydrogenase), and *putP* (proline permease) genes were excluded as potential targets due to their high rates of horizontal gene transfer; the *rrs* (16S rRNA) gene was excluded as a target due to the presence of multiple gene copies, with different sequences in a single genome. Based on the above criteria, the *mdh* (malate dehydrogenase) gene was selected as a target for development of a microbial source tracking assay. The *mdh* assay was optimized to analyze a 150 bp fragment corresponding to residues G191 to R240 (helices H10 and H11) of the Mdh catalytic domain. 295 fecal isolates (52 horse, 50 deer, 72 dog, 52 seagull and 69 human isolates) were sequenced and analyzed. Target DNA sequences for isolates from horse, dog plus deer, and seagull formed identifiable groupings. Sequences from human isolates, aside from a low level (ca. 15%) human specific sequence, did not group; nevertheless, other hosts could be distinguished from human. Positive and negative predictive values for two- and three-way host comparisons ranged from 60% to 90% depending on the focus host. False positive rates were below 10%. Multiple *E. coli* isolates from individual fecal samples exhibited high levels of sequence homogeneity, i.e. typically only one to two *mdh* sequences were observed per up to five *E. coli* isolates from a single fecal sample. Among all isolates sequenced from fecal samples from each host, sequence homogeneity decreased in the following order: horse>dog>deer>human and gull. For in-library isolates, blind analysis of fecal isolates ($n=12$) from four hosts known to contain host specific target sequences was 100% accurate and 100% reproducible for both DNA sequence and host identification. For blind analysis of non-library isolates, 18/19 isolates (94.7%) matched one or more library sequences for the corresponding host. Ten of eleven geographical outlier fecal isolates from Florida had *mdh* sequences that were identical to in-library sequences for the corresponding host from California. The *mdh* assay was successfully applied to environmental isolates from an underground telephone vault in California, with 4 of 5 isolates matching sequences in the *mdh* library. 146 sequences of the 645 bp *mdh* fragment from five host sources were translated into protein sequence and aligned. Seven unique Mdh protein sequences, which contained eight polymorphic sites, were identified. Six of the polymorphic sites were in the NAD+ binding domain and two were in the catalytic domain. All of the polymorphic sites

* Corresponding author. Tel.: +1 415 514 0102; fax: +1 415 476 7974.
E-mail address: Kathryn.Ivanetich@ucsf.edu (K.M. Ivanetich).

0167-7012/$ - see front matter
doi:10.1016/j.mimet.2006.04.026

were located in surface exposed regions of the protein. None of the non-silent mutations of the Mdh protein were in the 150 bp *mdh* target. The advantages and disadvantages of the assay compared to established source tracking methods are discussed.

Keywords: DNA sequencing; *Escherichia coli*; Fecal pollution; Indicator organisms; Malate dehydrogenase; *mdh*; Microbial source tracking; Sub-typing; Water quality

1. Introduction

The quality and safety of watersheds nationwide is threatened by fecal pollution from human sources, domestic and farm animals, and wildlife (see e.g. NRDC, 2004). Watershed contamination by fecal bacteria is associated with a wide variety of health hazards, including gastrointestinal and viral infections (Cabelli, 1977; Cabelli et al., 1982; Cabelli, 1983; Dufour, 1984; Dufour and Ballentine, 1986; USEPA, 1986; Pruss, 1998; USEPA, 2003). Once elevated levels of watershed fecal pollution have been documented, particularly for non-point pollution sources, the host sources of the pollution must be identified in order to effectively assess health risks and pursue remediation.

Identification of host sources of fecal pollution is typically achieved by microbial source tracking (MST), i.e. the differentiation of microorganisms, such as *Escherichia coli* or *Enterococcus*, on the basis of the host source (see Sinton et al., 1998; Scott et al., 2002; Simpson et al., 2002; Meays et al., 2004). MST methods are phenotypic or genotypic and can be library-based or library-independent. Examples include antibiotic resistance analysis, carbon source utilization, pulsed-field gel electrophoresis, repetitive element PCR, host-specific PCR, DNA ribotyping, and human pathogen analysis (Griffith et al., 2003; Harwood et al., 2003; Myoda et al., 2003; Noble et al., 2003). Although many of the aforementioned methods can successfully distinguish human from non-human fecal pollution, most can not accurately and reproducibly distinguish individual host sources of pollution (Griffith et al., 2003; Noble et al., 2003). Library-dependent methods often identified the main source of fecal pollution in blind inoculated water samples, but had high rates of false positives (Griffith et al., 2003; Myoda et al., 2003). Among library-based methods, genotypic methods generally performed better than phenotypic methods (Griffith et al., 2003).

Criteria for sub-typing of microbial organisms by DNA sequencing (Olive and Bean, 1999) do not appear to have been previously applied for identification of gene targets for library-dependent MST assays. These criteria are as follows: The target sequence must consist of a variable region flanked by highly conserved regions and must not be vulnerable to horizontal gene transfer. In addition, the variable region must be relatively short, and contain sufficient allelic polymorphism to differentiate strains (Olive and Bean, 1999).

Automated fluorescent DNA sequencing technology has numerous advantages for analysis of gene targets for MST assays. This technology can generate target gene sequences rapidly, accurately (>98.5% accuracy) and reproducibly. DNA sequencing technology is automated, inexpensive and widely available, and has recently been applied to MST (Ram et al., 2004).

E. coli was chosen as the target organism for development of an MST assay since this microorganism resides in the intestines of humans, other warm-blooded animals, and birds (Geldreich, 1966; Orskov and Orskov, 1981), is a widely used indicator of fecal pollution and the target organism for numerous MST assays, and has a fully sequenced genome.

The *E. coli* 936 bp malate dehydrogenase (*mdh*) gene was chosen as the target gene based on the criteria indicated above. The Mdh enzyme is a component of the citric acid cycle, and is composed of an NAD^+ binding domain (AA 1–150) and a catalytic domain (AA 151–312) (Hall et al., 1992). DNA sequencing of the *mdh* gene target was optimized and applied to 295 fecal *E. coli* isolates from five host species, namely horse, human, dog, deer and seagull. These hosts are the most likely sources of pollution in the San Pedro Creek and San Francisco Bay Watersheds, the first watersheds to which the assay will be applied. Application of the *mdh* assay to library isolates resulted in the identification of host specific sequences. For horse, seagull, the deer/dog pair, and in some cases human hosts, the *mdh* gene target sequence is capable of distinguishing hosts with positive and negative classification rates of 58% to 94%, and false positive rates of <10%. In a blinded study, the assay correctly identified and classified host specific, in-library target sequences with 100% reproducibility and accuracy. The assay has been applied to environmental isolates and to blinded analysis of non-library isolates from target hosts, with high success rates. For a limited number of samples, the *mdh* target sequence exhibited minimal geographical diversity. The nature and locations of non-silent mutations in the Mdh protein have been assessed.

2. Materials and methods

2.1. Gene target selection

Gene targets were analyzed by reported criteria (Olive and Bean, 1999). DNA sequences were globally aligned and dendrograms were prepared using the neighbor-joining algorithm of the CLUSTAL-X program (Higgins et al., 1996; Thompson et al., 1997). Bootstrap analysis was performed and indicated in dendrograms if the bootstrap percentage from 1000 iterations was >50%.

2.2. Primer analysis and design

Initially, reported *mdh* primers (Boyd et al., 1994) were used as PCR and external DNA sequencing primers for the *mdh* gene. These primers were subsequently analyzed using the published GenBank *E. coli* K12 *mdh* gene sequence and *mdh* sequences generated during preliminary investigations, with Primer Express software (Applied Biosystems, Foster City, CA) as described previously (O'Shaughnessy et al., 2003). PCR/external sequencing primers and internal sequencing primers for *mdh* were designed de novo by this approach. Primers were synthesized and purified as reported earlier (Ivanetich et al., 1999), and diluted to 50 μM with dH_2O.

2.3. Fecal sample collection and E. coli isolation, culture and archiving

Between 15 and 20 fecal samples were collected from each of the following hosts: humans, dogs, horses, deer and seagulls. For the first four hosts, fecal samples were collected in San Pedro Creek Watershed, Pacifica, California. For human, dog and horse hosts, each fecal sample was from a single individual. For deer, an effort was made to collect fecal samples from individual animals but mixed fecal samples were possible. For seagulls, combined fecal samples were collected from flocks in Half Moon Bay, California (the next significant watershed to the south of Pacifica). Samples were collected in these areas to support MST studies on the San Pedro Creek and San Francisco Bay Watershed. All samples were collected in Carry–Blair single-swab fecal collection tubes and refrigerated until processing. Samples were streaked on ECC Chromagar plates (Hardy Diagnostics, Santa Maria, CA) and incubated overnight at 37 °C. *E. coli* colonies were selected from other coliform bacteria based on a color indicator in ECC Chromagar. Five *E. coli* colonies from each plate were selected, visually checked for purity using a dissecting microscope, re-streaked on fresh ECC Chromagar plates, and incubated overnight at 37 °C. Individual *E. coli* colonies were confirmed with the Spot Indole test. The original mixed colony plate and the five *E. coli* single-colony plates were refrigerated for up to 3 days. A stab from each single colony *E. coli* plate was incubated overnight at 37 °C in LB broth without antibiotics. After incubation, 0.8 mL of culture was vortex mixed with 0.2 mL anhydrous glycerol (final concentration, 20% glycerol) and stored at −80 °C.

2.4. Genomic DNA purification and PCR amplification

Glycerol stocks of *E. coli* colonies were streaked onto LB plates without antibiotics (University of California San Francisco Cell Culture Facility), using a Difco brand 1 μL inoculating loop (Fisher Scientific, Cat. No. 22-031-20), and the plates were incubated at 37 °C overnight. A stab from each plate was inserted into 2 mL of LB in a 15 mL Falcon tube, capped loosely and incubated overnight at 37 °C. Genomic DNA was purified from 500 μL of *E. coli* culture with the MasterPure Complete DNA and RNA Purification Kit (Epicentre, Madison, WI). The protocol for cell samples was followed, with the exception that purified DNA pellets were resuspended in 35 μL dH_2O. Genomic DNA was stored at −20 °C.

The initial target sequence was an 864 bp fragment of the *mdh* gene amplified by published PCR primers (Boyd et al., 1994). In subsequent experiments, an 825 bp *mdh* gene fragment was amplified with the newly designed PCR primers. PCR reaction mixtures for amplification of the *mdh* gene from *E. coli* genomic DNA were prepared according to the Applied Biosystems AmpliTaq Gold Handbook 'Protocol for Amplification of Samples' and cycled on an MJ Research PTC-225 thermal cycler (Waltham, MA) using the "Touchdown" protocol (Don et al., 1991). Excess primers and dNTPs were removed with the QIAquick PCR purification kit (Qiagen, Valencia, CA). Aliquots of purified PCR products were run on 1% agarose gel and visualized with a BioDoc-It system (UVP, Upland, CA). The presence of the 825 bp or 864 bp *mdh* PCR product confirmed a successful PCR reaction. Purified PCR products were used for DNA sequencing and/or stored at −20 °C.

In one experiment the 825 bp *mdh* fragment was PCR amplified directly from glycerol stocks. Experimental conditions were as described in the methods, except that 1 μL of glycerol stock replaced the same volume of purified genomic DNA in the PCR reaction mixture.

2.5. DNA sequencing and data analysis

Sequencing reactions contained 1 μL purified PCR product, 3.5 pmol sequencing primer, and BigDye Terminator Master Mix v. 3.1 (3 μL of 2/3X, initial concentration) in 10 μL total volume and were subjected to 30 cycles, purified and sequenced on an Applied Biosystems PRISM® 3700 capillary sequencer with POP6 polymer (Foster City, CA) as described earlier (O'Shaughnessy et al., 2003). Initially, the 825 bp *mdh* PCR products were sequenced with double coverage, edited and assembled in ContigExpress (Invitrogen, Carlsbad, CA), and trimmed to 645 bp. In the optimized assay, a 394 bp *mdh* fragment was sequenced with double coverage, and the sequences were edited, assembled and trimmed to 150 bp (Table 1 and Fig. 1).

Average Phred scores and 1st Phred <20 scores, used to assess the quality of all DNA sequencing runs, were generated automatically by dnaLIMS software (dnaTools Inc., Ft. Collins, CO). The Average Phred score is calculated over the entire sequencing run, and the 1st Phred <20 score is the first base that the Phred score drops below 20. Average Phred scores are reported, but comparable results were obtained with 1st Phred <20 scores (Data not shown). Sequences with Average Phred scores above 30 for external primers (Primers 1 and 2) or above 22 for internal primers (Primers 3 and 4) were considered acceptable. Samples generating sequences with Average Phred scores below those values were re-sequenced. Unless otherwise indicated, reported values are means and standard deviations and statistical analysis was performed with the Student's *t*-test for single tailed *p* values. A significant difference between means was $p<0.01$.

Consensus sequences were generated and exported in FASTA sequence format. The sequences of the trimmed 150 bp region corresponding to G191 to R240 of the Mdh catalytic domain, or in preliminary experiments, the 645 bp region corresponding to Mdh residues S26 to R240, were globally aligned and dendrograms were prepared using the neighbor-joining algorithm of the CLUSTAL-X program (Higgins et al., 1996; Thompson et al., 1997). Abbreviations for host sources are as follows: dg, dog; dr, deer; gu, seagull; ho, horse; and hu, human.

From the initial dendrograms, redundant (identical) sequences from an individual fecal sample were removed from the multiple sequence alignments for clarity. To account for multiple identical sequences from an individual fecal sample, the *organism.sample.colony*×*number of identical sequences* format was used in the dendrograms. For example, if *mdh* sequences from five *E. coli* colonies isolated from dog #5 (dg5.1, dg5.2, dg5.3, dg5.4 and dg5.5) were identical, then *dg5.1.*×*5* was used to represent the five sequences in final multiple sequence alignments and dendrograms. The non-redundant sequences were again subjected to multiple sequence alignment by the CLUSTAL-X program using the neighbor-joining algorithm. The *mdh* sequence from *E. coli* O157:H7 (GenBank accession number: BA000007, bp 4118567–4119505, minus strand) was trimmed and added to the sequence alignment as an outgroup for the final dendrograms. Bootstrap analysis was performed and indicated in dendrograms if the bootstrap percentage from 1000 iterations was >50%.

Several parameters were calculated to assess the confidence of host sequence clustering patterns in the dendrograms (Myoda et al., 2003). Sensitivity (true positive rate) is the likelihood that a sequence belonging to a given host species has a positive test result. Positive predictive value is the likelihood that a positive test result is true for the focus species. Specificity (true negative rate) is the likelihood that a sequence not belonging to a given host species has a negative test result. Negative predictive value is the likelihood that a negative test result is correct. Test efficiency is the likelihood that any sequence was classified correctly. The false positive rate is the percentage of sequences that were classified incorrectly, based on the focus species of the comparison.

Table 1
PCR and DNA sequencing primers designed for the *mdh* gene

Primer	Primer usage	Sequence (5′–3′)	T_m (°C)	Location on *mdh* gene
1	PCR and sequencing	TGAAAGTCGCAGTCCTCGG	59	644–663
2	PCR and sequencing	TCCACGCCGTTTTTACCC	58	1469–1452
3	Sequencing	GGCGTTACCACGCTGGG	53	1075–1091
4	Sequencing	GCACTTCAACTTCGCCTGG	56	1137–1156

In the optimized assay, the 825 bp *mdh* gene fragment (bases 644–1469) is PCR amplified with primers 1 and 2, and the 394 bp *mdh* fragment (bases 1075–1469) is sequenced with primers 2 and 3.

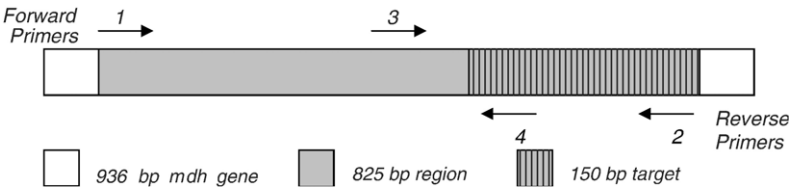


Fig. 1. Schematic of the 936 bp *mdh* gene including the 825 bp PCR product (gray including gray striped area) and the 150 bp segment of the catalytic domain (black stripe) sequenced and analyzed in the optimized assay.

2.6. Analysis of Mdh protein sequence polymorphisms

The DNA sequences of the 645 bp *mdh* fragment from 146 isolates (16 dog, 43 deer, 28 seagull, 22 horse and 37 human isolates) plus the *E. coli* O157:H7 outgroup sequence were translated using the Transeq program in the EMBOSS suite (http://www.ebi.ac.uk/emboss/ transeq/). Proteins were translated in the first reading frame and saved in FASTA format. Mdh protein sequences were grouped by host source and aligned. Host specific and subsequent alignments were performed with the T-Coffee web server (http://igs-server.cnrs-mrs.fr/Tcoffee/ tcoffee_cgi/index.cgi) (Poirot et al., 2004) using the advanced options, with all settings at default parameters. In each host-specific Mdh protein sequence alignment, non-redundant sequences were identified. After the Mdh sequences for each host were aligned individually, the non-redundant sequences from each host source were pooled and aligned, and redundant sequences were removed. Eight non-redundant sequences were identified, and subsequently aligned with full length Mdh protein sequences from structures in the protein data bank (1EMD, *E. coli* Mdh protein complexed with citrate and NAD, and 2CMD, *E. coli* Mdh protein complexed with citrate) and with the Mdh sequence translated from the *E. coli* O157:H7 *mdh* gene sequence in order to identify the amino acid polymorphisms in the 645, 495 and 150 bp *mdh* gene fragments. All three-dimensional structure analysis of Mdh proteins was performed using University of California, San Francisco Chimera: An Extensible Molecular Modeling System (http://www.cgl.ucsf.edu/chimera/) (Pettersen et al., 2004).

3. Results

3.1. Selection of target gene

The *16S* rRNA gene (*rrs*) of *E. coli* was excluded as a potential gene target for a DNA sequencing MST assay based on several factors. First and most important, there are seven copies of the *rrn* operon (*rrn A, B, C, D, E, G,* and *H*) in a single *E. coli* genome, and inter-cistronic heterogeneity among the seven *rrs* genes would produce multiple distinct sequences for any single *E. coli* isolate. This heterogeneity among the *rrs* genes of one *E. coli* isolate renders DNA sequencing analysis extremely problematic since DNA sequence may not be determined accurately and unambiguously when sequence heterogeneity exists. Although not included in the criteria for sub-typing of microbiological organisms (Olive and Bean, 1999), either a single gene copy or multiple identical copies of a single gene within the genome is essential for accurate, unambiguous DNA sequencing, including sequencing of gene targets for MST. Second, if one intends to compare a selected *rrs* sequence among isolates, one must choose carefully: For example, the *rrsG* of *E. coli* O157:H7 shares higher sequence identity with the *rrsA*, *rrsB*, and *rrsE* of *E. coli* K-12 MG1655 than with the *rrsG* of *E. coli* K-12 MG1655 (Fig. 2). Since differences among inter-cistronic copies *within* one genome may be greater than those among sequences of the target gene derived from distinctly different genomes, phylogenetic trees constructed from the comparison of sequences of the rrsG and other ostensibly representative 16S rRNA genes among different strains may not be productive. Third, it has been reported that the *16S* rRNA gene is not sufficiently discriminative and that there is no correlation between *rrs* sequence and host source (Guan et al., 2002).

Several additional genes were identified as potential targets for MST, and examined by the criteria for sub-typing of microbial organisms by DNA sequencing (Olive and Bean, 1999). They include the *icd, putP* and *mdh* genes. The *icd* gene encodes the citric acid cycle enzyme isocitrate dehydrogenase, and like the *mdh* gene described below, was among the 11 loci used to establish the *E. coli* standard reference strains (Ochman and Selander, 1984). The *icd* locus appears to possesses high allelic diversity: According to a study of the genetic structure of *E. coli* from Australian mammals, not only does the *icd* locus possess more alleles than the *mdh* locus, the frequencies of *icd* alleles are also more evenly distributed (Gordon and Lee, 1999). However, frequent

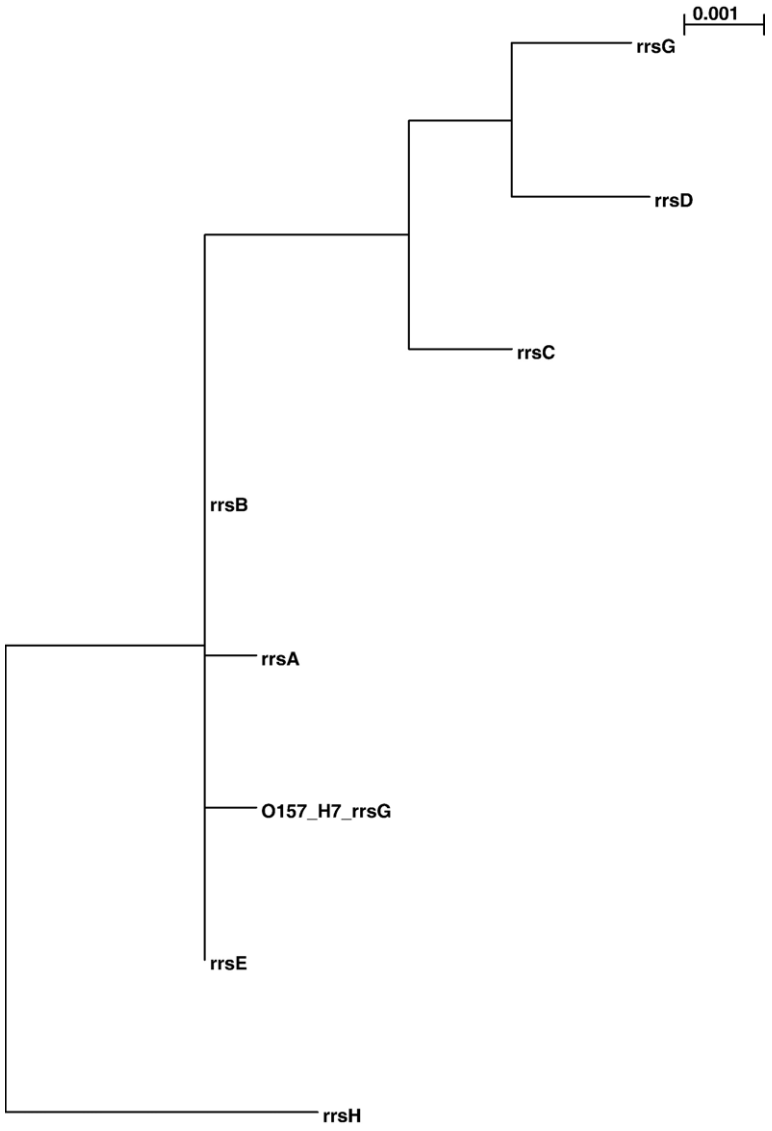


Fig. 2. Dendrogram constructed from seven *rrs* genes of *E. coli* K-12 MG1655 and one *rrsG* of *E. coli* O157:H7. The NCBI accession number of K-12 MG1655 and *E. coli* O157:H7 genomes are U00096 and BA000007, respectively.

horizontal gene transfers at the *icd* locus are at least partially responsible for its higher allelic diversity, and these events compromise the usefulness of *icd* for MST. For the 1251-bp *icd* gene, a crossover point at bp 1087 by many strains of lambdoid phage 21 and another at bp 1,098 by defective prophage element e14 have been reported (Blattner et al., 1997; Wang et al., 1997a,b). The 216-bp *icd* replacement segment of phage 21 was found in 39% of *E. coli* strains, including the standard strains ECOR 38 and ECOR 39 (Wang et al., 1997a,b). In the *E. coli* K-12 MG1655 genome, the 3′ end of the *icd* locus spanning bp 1098 to bp 1251 has been attributed to the invasion of defective prophage element e14 (Blattner et al., 1997). The existence of recombination events among *E. coli* strains at the *icd* locus will produce phylogenetic trees completely inconsistent with the true evolutionary tree.

Although the genetic diversity at the *icd* locus is very high in the vicinity of the bp 1087 crossover point based on a multiple sequence alignment of the *icd* locus for EC10, 14, 15, 17, 32, 37, 40, 52, 58, 64, 69, and 70 and K-12 (Multiple sequence alignment not shown), if one removes the 3′ end of the *icd* locus which is subject to crossover, the allelic polymorphism rate decreases significantly in the remaining sequence. In addition, the insertion sequence of phage 21 comprising the 165-

bp *icd* replacement segment, the downstream 1143-bp integrase gene *int*, and an intermediate 113 bp segment between them, render the design of PCR and DNA sequencing primers very difficult.

Upon initial examination, the *putP* (proline permease) gene appeared to be a promising gene target candidate: It contains 108 polymorphic nucleotide sites in the 1467-bp partial gene (Nelson and Selander, 1992) (comprising 97% of the full gene). On average, the sequences of pairs differed at 2.4% of nucleotide sites for the *putP* locus (Nelson and Selander, 1992), but only 1.1% at the *mdh* locus (Boyd et al., 1994), suggesting that the *putP* locus may contain more useful sites for microbial subtyping than the *mdh* gene. Furthermore, since the phylogenetic tree for the *putP* was generally congruent with a tree based on multilocus enzyme electrophoresis (MLEE) and a tree for the *gapA* gene, which encodes glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase and during the 1990s was thought not susceptible to recombination events (Nelson et al., 1991; Nelson and Selander, 1992), it appeared that horizontal gene transfer at the *putP* locus is not significant.

However, Nelson and Selander's finding of no significant horizontal gene transfer at *putP* (Nelson and Selander, 1992) has been challenged. The phylogenetic tree for *putP* shows significant incongruence with the tree for whole-genome data which was based on MLEE and random amplified polymorphic DNA (RAPD) (Lecointre et al., 1998). In addition, the molecular divergence for *gapA* is smaller than for other loci, suggesting a recent evolutionary event which has purged most of the variability from the *E. coli gapA* locus, rendering the basis for Nelson and Selander's conclusions about *putP* (Nelson and Selander, 1992) untenable. Finally, Escobar-Páramo et al. (2004) recently confirmed that the *putP* locus is susceptible to horizontal gene transfer.

The 936 bp *mdh* gene of *E. coli* satisfied the criteria for typing microbial organisms by DNA sequencing (Olive and Bean, 1999) and was selected as gene target for development of a DNA sequencing MST assay based on the following: First, the 936 bp *mdh* gene is a sufficiently short target. Second, the *mdh* gene has 40 allelic polymorphisms among 19 ECOR reference strains from five major evolutionary lineages of *E. coli* and one strain not assigned to an ECOR group, which may provide sufficient sequence diversity for MST. Third, horizontal gene transfer in the vicinity of the *mdh* locus is a rare event as confirmed by multi-locus enzyme electrophoresis of the 72 *E. coli* strains that comprise the ECOR standard reference collection, gene-level evolutionary trees and the evolutionary relationships among pathogenic and non-pathogenic *E. coli* strains (Ochman and Selander, 1984; Boyd et al., 1994; Pupo et al., 1997). Finally, there was no evidence for more than a single copy of the *mdh* gene in the *E. coli* genome.

3.2. DNA sequencing of the mdh gene target

In initial experiments, an 864 bp *mdh* gene fragment was PCR amplified and sequenced with published primers (Boyd et al., 1994). Average Phred scores were 12±5 and 16±10, and coefficients of variance (CVs) were 46% and 63% (n=27 for each primer), indicative of failed or poor quality sequencing runs. The published *mdh* primers were subsequently analyzed and found to have T_ms of 68 °C and 75 °C, which are well above the optimal T_m range (45 °C to 60 °C) for automated sequencing (Applied Biosystems, 2000; Applied Biosystems, 2001). In addition, the published *mdh* primers were 33 to 35 bases in length, far longer than the optimal primer size of 17 to 24 bases for automated sequencing (Applied Biosystems, 2000; Applied Biosystems, 2001). Furthermore, the published 3′ *mdh* primer had a three-base mismatch to the published GenBank *E. coli* K12 *mdh* gene sequence. New *mdh* primers were designed de novo to comply with fluorescent dye terminator DNA sequencing guidelines. Optimized *mdh* primers 1 and 2 were PCR and external sequencing primers, and *mdh* primers 3 and 4 were internal sequencing primers (Table 1 and Fig. 1). The optimized primers had no mismatches to the published *E. coli* K12 *mdh* sequence, had optimal lengths and T_m values, and PCR amplified an 825 bp *mdh* gene fragment.

PCR amplification and sequencing with *mdh* primers 1 and 2 resulted in high quality sequencing runs, characterized by Average Phred scores of 35±16 and 38±14 and CVs of 38% to 45% (n=33 for each primer). The Average Phred scores were significantly increased and the CVs were significantly decreased compared to PCR amplification and sequencing of the 864 bp *mdh* fragment with the published primers (p<0.0005).

Internal sequencing with optimized primers 3 and 4 of the 864 bp *mdh* PCR product generated from the published primers resulted in Average Phred scores of 25±13 and 24±12 (n=16 for each primer) and CVs of 50%. In contrast, sequencing the 825 bp *mdh* PCR product with internal primers 3 and 4 produced higher quality sequencing results. For primer 4, the Average Phred score was 33±9 and the CV was 26% (n=33), which was significantly improved (p<0.005) relative to sequencing the 864 bp PCR product. For primer 3, the Average Phred score was 25±7 (n=33), which was

comparable to that for sequencing the 864 bp PCR product, but reproducibility improved significantly, with a CV of 28%, i.e. approximately half that for sequencing the 864 bp PCR product generated from the published primers.

3.3. Refinement of the mdh gene target

The trimmed 645 bp *mdh* gene fragment generated from the 825 bp PCR fragment was divided into a 495 bp fragment corresponding to Mdh residues S26 to P190 and a 150 bp catalytic domains fragment corresponding to Mdh residues G191 to R240 (helices H10 and H11). 146 sequences from 16 dog, 43 deer, 28 seagull, 22 horse and 37 human isolates, plus the *E. coli* O157:H7 outgroup sequence, were analyzed. Excluding the outgroup, the 495 bp *mdh* fragment contained 25 polymorphic sites. However, many of the polymorphic sites in this region were found in no more than two isolates and would not be useful for categorizing host isolates. In contrast, the 150 bp fragment contained only 10 polymorphic sites, but exhibited a 30% higher density of polymorphic sites than the 495 bp fragment. Each of the polymorphic sites in the 150 bp fragment was found in multiple isolates.

A subset of sequences of the trimmed 645 bp *mdh* gene fragment from two hosts, i.e. 16 dog and 23 horse sequences, was subjected to multiple sequence alignment. The derived dendrogram distinguished the two hosts (Fig. 3). In addition, differences between sequences within each set were observed, e.g. ho12.4, ho13.2 and ho17.1 differed slightly from other sequences in the horse set (Fig. 3). When the target was shortened to a 282 bp or 150 bp fragment corresponding to Mdh residues T147 to R240 or G191 to R240, the boundary of the two sets was preserved in the dendrogram (Fig. 4). The topology of the dendrograms generated from the 282 bp and 150 bp *mdh* sequences were identical (Fig. 4). However, shortening the *mdh* target sequence from 645 bp to either 282 bp or 150 bp resulted in three dog sequences in the gray box of Fig. 3 moving into the horse subset to become false positives and the loss of nuances within the target sequences for each host set, i.e. a single target sequence for each host (Fig. 4). Since the reduction of target sequence length by four-fold did not significantly sacrifice the sensitivity and specificity of the dendrograms, but decreased the number of sequencing reactions by 50%, the 150 bp *mdh* catalytic domain fragment corresponding to Mdh residues G191 to R240 was chosen as the target sequence for further sequencing and analysis of fecal samples from multiple host species.

3.4. Host specific sequence differences in the 150 bp mdh catalytic domain fragment

The sequences of the 150 bp *mdh* catalytic domain were obtained for 295 isolates (72 dog, 50 deer, 52 seagull, 52 horse and 69 human), and a multiple sequence alignment including the *E. coli* O157:H7 published sequence was constructed using the CLUSTAL-X program. Excluding the *E. coli* O157:H7 outgroup (not shown), 10 polymorphic sites were identified in the multiple sequence alignment and were used to discriminate host species (Fig. 5). For each of the five host species, sequence variation in the 150 bp *mdh* fragment was assessed for multiple *E. coli* isolates from individual fecal samples. For human, horse and dog hosts, where each fecal sample was from a single individual, 69% to 80% of fecal samples contained a single *mdh* sequence and 13% to 27% of fecal samples contained two sequences among 3.4 ± 1.2 to 4.6 ± 0.6 isolates sequenced per fecal sample. For these hosts, 93% to 100% of the fecal samples contained one or two *mdh* sequences. For deer, where an effort was made to collect fecal samples from individual animals but where mixed fecal samples were possible, 40% of the fecal samples had a single *mdh* sequence and 53% had two *mdh* sequences in 3.3 ± 1.3 isolates sequenced per fecal sample. Thus, 93% of deer fecal samples contained one or two *mdh* sequences per approximately 3 isolates. For the human, horse, dog, and deer host species, three or more *mdh* sequences were found in $\leq 7\%$ of the individual fecal samples. In contrast, for mixed fecal samples from seagull flocks, 29% and 36% of the fecal samples exhibited one and two *mdh* sequences, respectively, and 36% of the fecal samples had three or four different *mdh* sequences per fecal sample, with 4.1 ± 1.0 isolates analyzed per fecal sample. The degree of sequence homogeneity among all fecal isolates from an individual host varied by host species. Based on the percentage of fecal isolates with sequences falling within the most populated sequence group, sequence homogeneity decreased in the following order: horse > dog > deer > human and seagull. For horse, 39 of 50 isolates (78%) had one target sequence. For dog, 44 of 72 isolates (61%) and for deer, 22 of 50 isolates (44%) were in the single most populated sequence group. For human and seagull, 19 of 69 isolates (28%) and 15 of 57 isolates (26%) fell in the most populated sequence group.

Dendrograms for two and three way host comparisons, i.e. horse vs. dog, horse vs. human, horse vs. dog and seagull, and dog vs. deer, for the 150 bp *mdh* catalytic domain fragment are shown in Figs. 4, 6 and 7. The sequences of the 150 bp *mdh* fragment from horse,

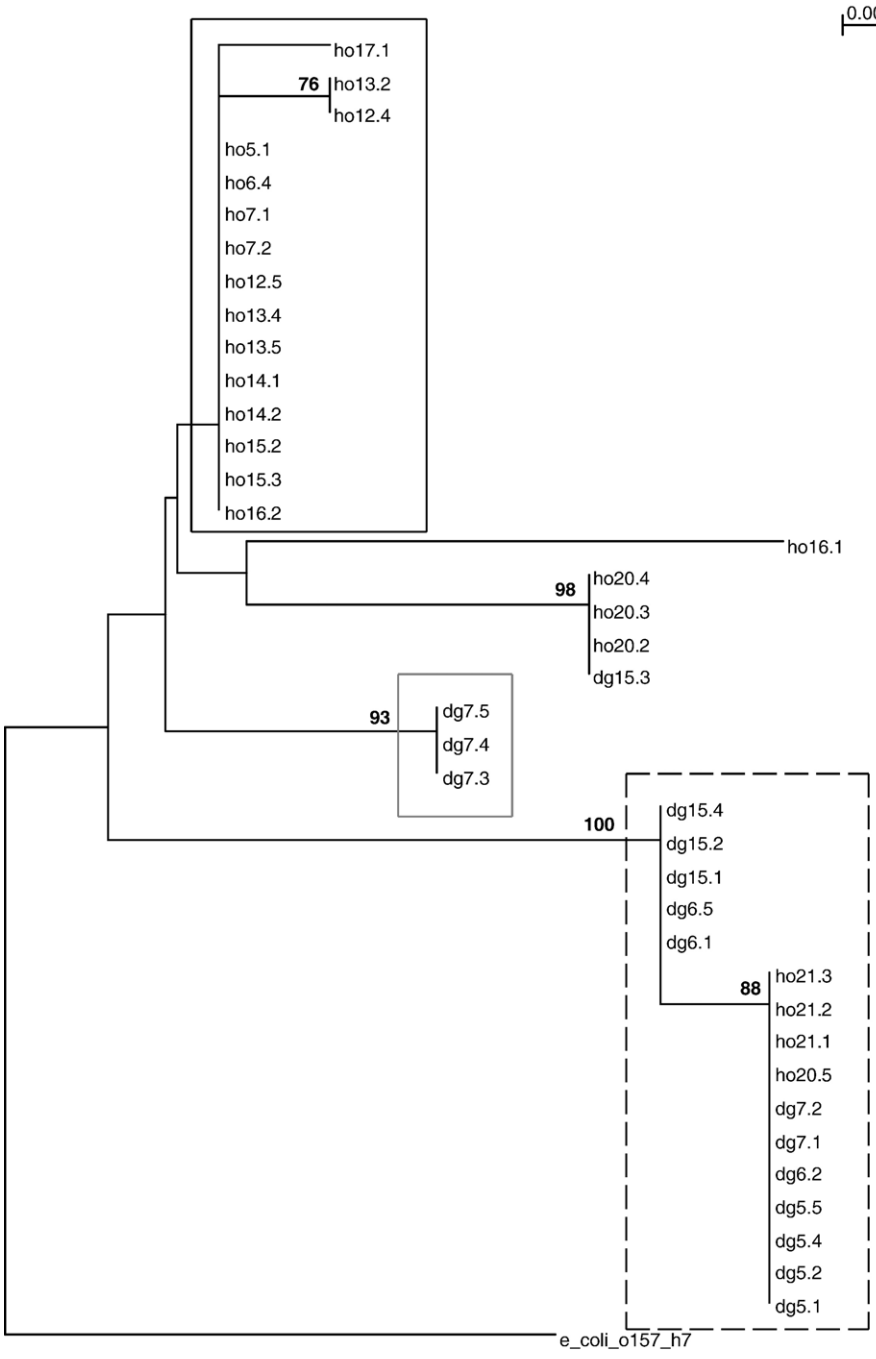


Fig. 3. Dendrogram based on 645 bp sequence which codes for Mdh residues S26 to R240 from 16 dog and 23 horse sequences. The target sequences for horse are in the solid rectangle, and for dog are in the dashed rectangle. Dog sequences in the gray box will move into the horse set if the target sequence is shortened to 282 or 150 bp. Numbers above branches are bootstrap percentages >50%.

seagull and dog each clustered into identifiable sequence groups indicated by boxes (Fig. 6). The human sequences showed extensive diversity, and with the exception of a low incidence sequence comprising approximately 15% of the total human sequences, did not form an identifiable group. Nevertheless, the other four hosts could be distinguished from human based on *mdh* target sequences. The parameters calculated for two and three way host comparisons are given in Table 2. Focus hosts were horse, dog and seagull, each of which

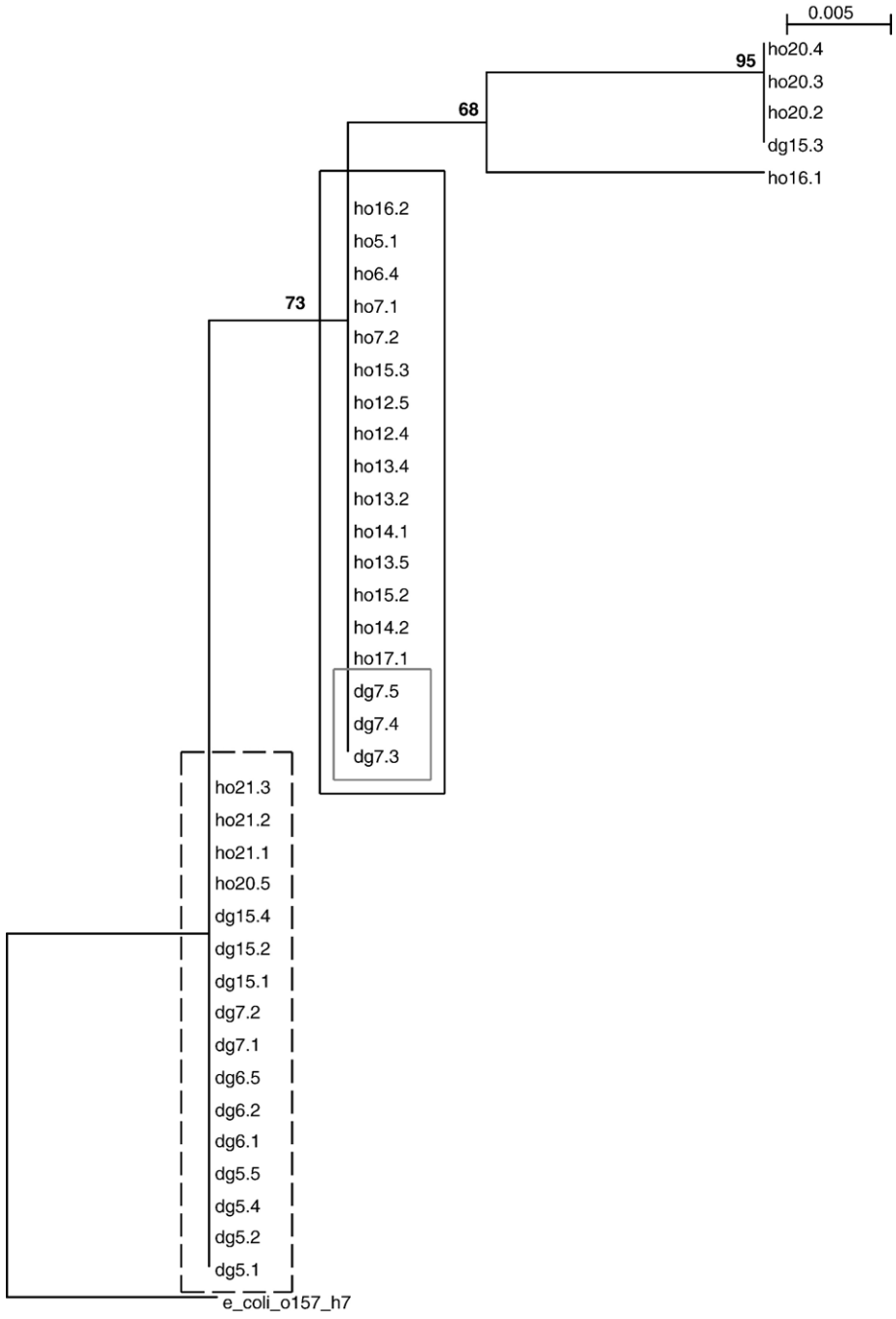


Fig. 4. Dendrogram based on the 282 bp *mdh* sequence (Mdh residues T147 to R240) for the set of sequences in Fig. 3. Numbers above branches are bootstrap percentages > 50%. An identical dendrogram was obtained for the 150 bp sequence coding for Mdh catalytic domain residues G191 to R240.

had a single target sequence or a collection of target sequences. For horse as the focus host, in two way comparisons, positive predictive values and sensitivities of 75% to 85% and negative predictive values and specificities of 83% to 90% were obtained. Similar values were obtained for three way host comparisons with horse as the focus species, except that positive predictive value decreased to 68% and negative predictive value increased to 90%. For dog in two and three way host comparisons, positive predictive values, sensitivities and specificities were approximately 90%, 61% and 92%, respectively, and negative predictive

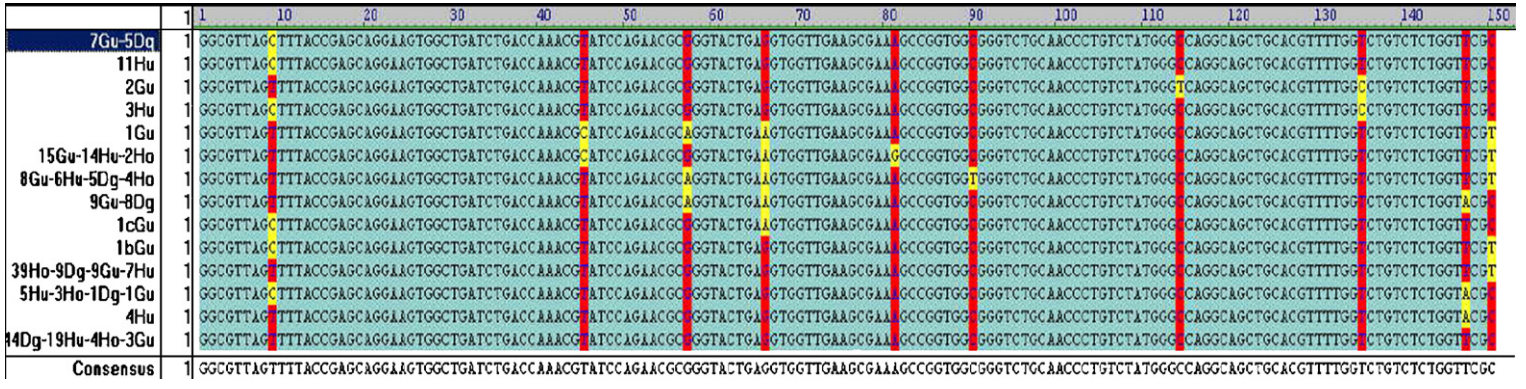


Fig. 5. Multiple sequence alignment of the 150 bp *mdh* catalytic domain fragment from 295 isolates, including 72 dog, 50 deer, 52 seagull, 52 horse and 69 human isolates, and one *E. coli* O157:H7 published sequence. Polymorphic sites are indicated in red and yellow. Numbers above branches are bootstrap percentages >50%.

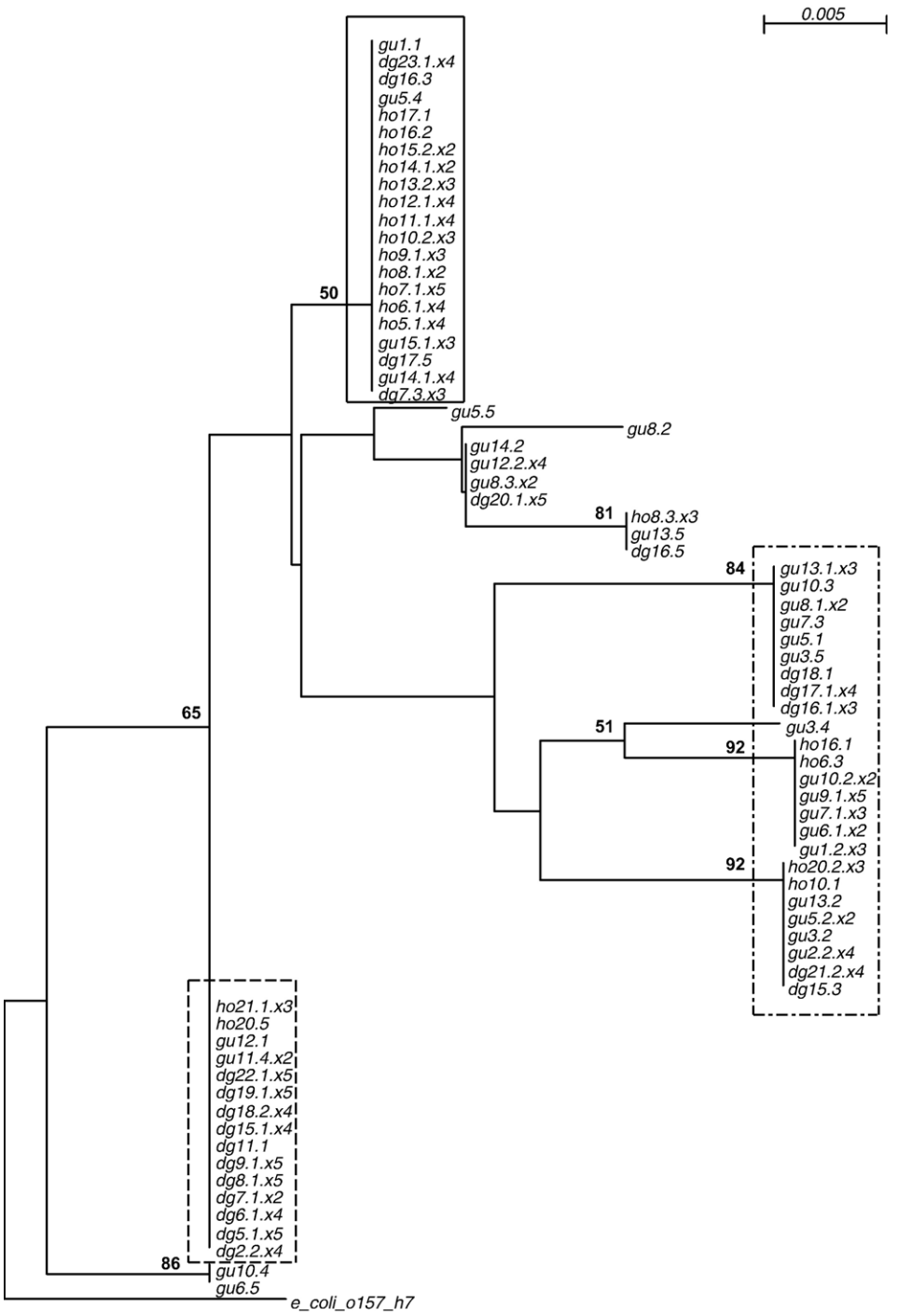


Fig. 6. Dendrogram constructed from 52 horse, 72 dog, and 50 seagull sequences of the 150 bp *mdh* catalytic domain fragment. The target sequence(s) for horse are in the solid rectangle, for dog in the dashed rectangle, and for seagull in the dash–dot–dash rectangle. Numbers above branches are bootstrap percentages > 50%.

values ranged from 63% to 78%. Seagull in a three-way host comparison was characterized by positive predictive value and sensitivity of ca. 60%, and negative predictive value and specificity of ca. 83%. For all two or three way host comparisons, test efficiencies ranged from 74% to 83% and false positive rates were from 3% to 10%. Based on dendrograms for the 645 bp or 150 bp *mdh* gene fragments, the sequences from dog and deer

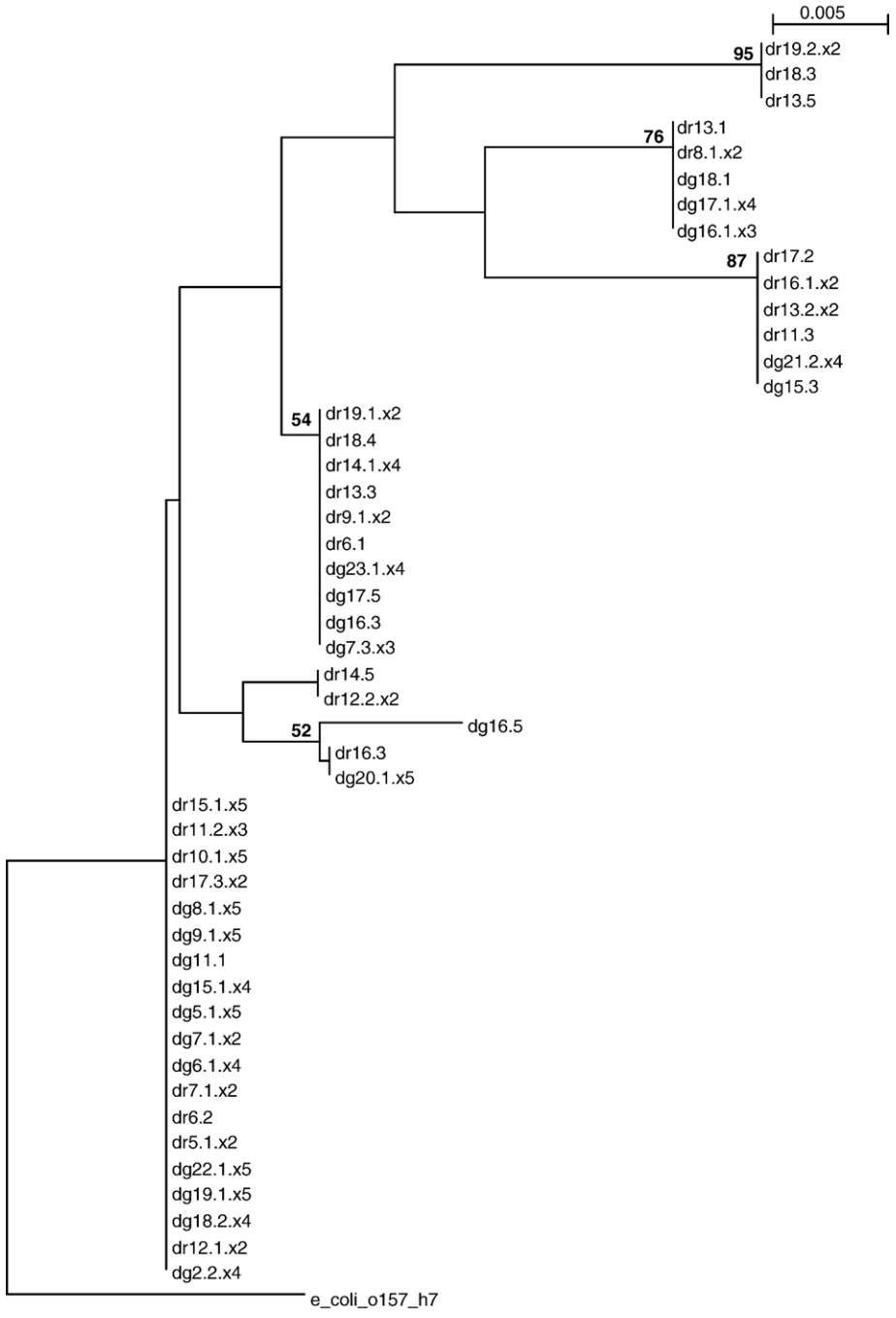


Fig. 7. Dendrogram constructed from 72 dog and 46 deer isolates from sequences of the 150 bp mdh catalytic domain fragment.

were extensively intermingled, and these hosts were indistinguishable (Fig. 7). If deer sequences replace dog sequences in a three way host comparison, i.e. deer or dog versus seagull and horse, the dendrogram shows three clusters, each dominated by one host species; this was observed for three way comparisons including either deer or dog isolates. To some extent, the deer, seagull, and horse dendrogram preserves host species boundaries as observed in the dog, seagull, and horse host comparison.

Table 2
Parameters calculated from dendrograms for two- and three-way host comparisons based on the 150 bp *mdh* gene target

Focus host	Compared to	Positive predictive value (%)	Sensitivity (%)	Negative predictive value (%)	Specificity (%)	Test efficiency (%)	False positive rate (%)
Horse	Dog	81	75	83	88	82	7
Horse	Human	85	75	83	90	83	6
Horse	Dog and seagull	68	75	90	86	83	10
Dog	Horse	92	61	63	92	74	3
Dog	Horse and seagull	86	61	78	94	81	4
Seagull	Dog and horse	63	58	81	85	76	10

3.5. Blind analysis of in-library isolates with host target sequences

Nine fecal *E. coli* isolates with the host target sequence (3 dog, 3 horse, and 3 seagull) were chosen from the library of isolates for a blind study. In addition, three in-library isolates with a human specific sequence that was found in 11 of 69 human isolates were assayed. Aliquots from glycerol stocks were blinded, cultured, and diluted in water, prior to isolation of genomic DNA. The sequences of the 150 bp *mdh* fragment from all 12 blinded isolates matched the known sequences of the isolates with 100% accuracy and reproducibility. All isolates were classified to the correct host source with 100% accuracy.

3.6. Geographic diversity

Ten of eleven (91%) *E. coli* isolates from fecal samples collected in Florida (3 dog, 5 deer, and 2 horse) had 150 bp *mdh* fragment sequences which were identical to sequences for the corresponding host from California. The sequence for one horse isolate did not match any sequences in the library. One additional dog isolate had insufficient sequence identity with the *E. coli mdh* gene, was presumed to be from a non-*E. coli* isolate, and was treated as an outlier.

3.7. Analysis of environmental samples

Five *E. coli* isolates from water pooling in an underground telephone company vault in San Mateo County, CA, characterized by elevated levels of *E. coli* from unknown host sources, were analyzed. Four of the five isolates were found to have 150 bp *mdh* fragment sequences that were identical to seagull target *mdh* sequences. The sequence of the remaining isolate did not match any sequences in the reference library.

3.8. PCR amplification from E. coli glycerol stocks

The 825 bp *mdh* gene fragment was PCR amplified directly from glycerol stocks for two *E. coli* isolates from each of five host species, which had provided excellent sequencing data with the standard protocol. For the standard protocol, which involves PCR amplification from purified genomic DNA, Average Phred scores were 49 ± 3 and 30 ± 2 for sequencing with primers 2 and 3, respectively. PCR amplification off glycerol stocks produced clean, strong bands on agarose gel for all isolates and generated Average Phred scores of 45 ± 3 and 25 ± 1 for primers 2 and 3, respectively. Since the Average Phred scores were 11% to 23% lower for PCR amplification directly from glycerol stocks ($p < 0.0001$, two tailed Students *t*-test for paired data), unless otherwise indicated, DNA sequencing was on PCR products amplified from purified DNA.

3.9. Blind analysis of non-library isolates

The *mdh* gene target was sequenced from DNA amplified directly from glycerol stocks of fecal *E. coli* isolates. Nine fecal *E. coli* isolates from dog and horse hosts from individuals not represented in the reference library, and 10 isolates from deer and seagull fecal samples collected separately from in-library samples, were analyzed. For 95% (18/19) of the isolates, i.e. 5 deer, 6 horse, 5 gull and 2 dog isolates, each isolate's sequence exactly matched one or more library sequences from the corresponding host. The sequence for the remaining isolate, from dog, matched a single library sequence for a seagull isolate.

3.10. Analysis of Mdh protein sequence polymorphisms

Translation of the 645 bp *mdh* sequences from 146 isolates from five host sources, alignment of Mdh protein

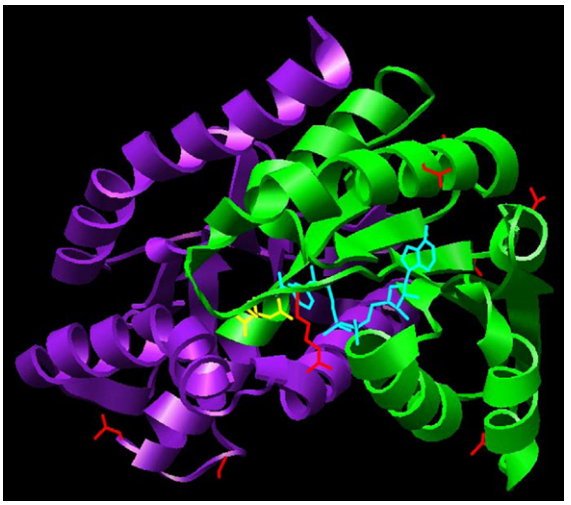


Fig. 8. Three-dimensional structure of Mdh from *E. coli* K-12, in complex with citrate and NAD+ cofactor (pdb ID: 1EMD). Side chains of polymorphic amino acid residues are depicted as wire models and colored red. The NAD+ binding domain and catalytic domains are depicted in ribbon form and colored green and purple, respectively. Bound NAD+ and citrate are represented as wire models, and colored cyan and yellow, respectively.

sequences for each host source, and pooling and alignment of non-redundant protein sequences from all hosts, resulted in the identification of eight non-redundant Mdh protein sequences. Of the eight unique protein sequences corresponding to the 645 bp gene fragment, only seven were unique when aligned with sequences of Mdh from the Protein Data Bank files 1EMD and 2CMD, and Mdh from *E. coli* O157:H7. The seven unique sequences contained eight polymorphic sites, identified as T51I, T64I, D71N, R80A, N100I, A106S, L161Q, and P166L. Six of the eight polymorphic sites were located in the NAD^+ binding domain (AA 1–150), whereas two sites were localized to the catalytic domain (AA 151–312). All eight of the non-silent mutations in the Mdh protein were in the 495 bp *mdh* gene fragment; none were in the 150 bp fragment. In addition to primarily being localized to the NAD^+ binding domain of Mdh, all of the polymorphic sites were located in surface exposed regions of the protein (Fig. 8).

4. Discussion

Most MST methods are library-based phenotypic or genotypic methods which attempt to characterize indicator microorganisms by host source, but typically do not directly analyze the gene sequence differences on which they are based. Examples include DNA ribotyping and antibiotic resistance analysis. Technologies that directly sample gene sequence, such as automated DNA sequencing, with few exceptions, have not been applied to MST. While this work was in progress, a DNA sequencing assay for the *E. coli uidA* gene that codes for β-glucuronidase was reported (Ram et al., 2004).

Although criteria for selection of gene targets for sub-typing of microbial organisms by DNA sequencing have been previously established (Olive and Bean, 1999), the present study provides the first example of the application of these criteria to select a gene target, i.e. the *E. coli mdh* gene coding for malate dehydrogenase, for the development of an MST assay, and to eliminate unsuitable targets, i.e. the *E. coli icd* gene coding for isocitrate dehydrogenase and the *putP* gene encoding proline permease, due to a significant rate of horizontal gene transfer. Although the ribosomal RNA gene is the classic target used to construct phylogenetic structures (Woese et al., 1990) and DNA ribotyping based on the 16S rRNA gene is a widely used MST method, the *E. coli rrs* gene coding for 16S rRNA was also eliminated as a suitable gene target for a DNA sequencing based MST assay, based on a criterion for typing by DNA sequencing developed by the authors, namely that successful gene targets should have a single gene copy in the genome or multiple copies of the target gene with identical sequences. This approach to gene target selection has facilitated cost-effective MST assay development by rational selection of promising gene targets and minimization of experimentation on unsuitable gene targets.

DNA sequence analysis of the *E. coli mdh* gene target for MST was optimized by re-design of PCR and DNA sequencing primers and refinement of the size of the target gene fragment. PCR and sequencing the 825 bp *mdh* fragment with optimized primers increased Average Phred scores by two- to three-fold compared to PCR amplification and sequencing of the 864 bp *mdh* gene fragment with published primers (Boyd et al., 1994). Primer optimization typically converted poor or failed sequencing runs (Average Phred scores ca. 15) to successful runs (Average Phred scores ca. 35), and significantly improved the reproducibility of the sequencing data, i.e. decreased the CVs for Phred scores by up to 40% (See Results). Optimization of the *mdh* target sequence length, i.e. the four-fold reduction of the sequence length of the trimmed *mdh* gene target from the 645 bp *mdh* fragment containing major portions of the NAD^+-binding and catalytic domains to the 150 bp *mdh* partial catalytic domain fragment (residues G190 to R240), decreased the number of sequencing reactions by 50% and streamlined sequence analysis without

significantly sacrificing the sensitivity and specificity of dendrograms (Figs. 3 and 4).

Based on analysis of 295 *E. coli* isolates from five hosts which potentially contribute significant fecal pollution to the first test watersheds, the 150 bp *mdh* target sequence was capable of distinguishing between selected hosts in two and three way host comparisons. For example, horse could be distinguished from dog or human, or from dog and seagull, with high rates of positive and negative predictivity and low false positive rates (Table 2). Dog could be distinguished from horse and seagull. However, dog and deer sequences co-mingled (Fig. 7) and could not be distinguished from each other, although deer was able to substitute for dog in the above host comparisons while substantially retaining host groupings. Although hosts such as horse, deer and dog can be distinguished from human by this assay, human sequences do not form an identifiable group, except for a low level (15% of total human isolates) human specific sequence. Thus, DNA sequencing of the 150 bp *mdh* catalytic domain fragment appears sufficient for the identification of horse, seagull and dog and/or deer and, in some cases, human fecal pollution among a limited range of hosts.

The *mdh* assay appears to provide the highest specificity for horse fecal pollution compared to other hosts, and provides comparable or improved rates for horse identification relative to other MST assays. For horse, the *mdh* target provided positive and negative predictivities of 68% to 90% for two and three host comparisons (Table 2). In comparison, ribotyping had a 92% rate of correct classification for horse in a three-host comparison and 49% to 61% rates of correct classification in eight-host comparisons, and rep-PCR had a 67% rate of correct classification for horse in an eight-host comparison (Carson et al., 2003). Similarly, the *mdh* MST assay had significantly better false positive rates for horse (6–10%) (Table 2) than several reported MST methods had across a wide variety of hosts (see below).

The *mdh* gene target sequence appears to provide equivalent or improved accuracy and selectivity in host identification (Table 2) compared to several other MST methods (Griffith et al., 2003; Myoda et al., 2003). For example, for two and three way host comparisons the *mdh* assay was characterized by positive predictivities ranging from 63% to 92% and specificities of 85–94%. In comparison, other methods had slightly less favorable ranges of values: rep PCR, ribotyping and pulse field gel electrophoresis were characterized by positive predictivities ranging from 38% to 86% and specificities of 50–100% and 0–67% (Myoda et al., 2003), while antibiotic resistance analysis, multiple antibiotic resistance and carbon source utilization provided positive predictivities ranging from 52% to 72% and specificities of 53–100% and 33–61% (Harwood et al., 2003). Sensitivities and negative predictivities for the *mdh* assay were comparable to the aforementioned methods (Harwood et al., 2003; Myoda et al., 2003). In addition, across several targets, the *mdh* assay had relatively low false positive rates (<10%), while several PCR methods, pulse field gel electrophoresis, antibiotic resistance analysis, multiple antibiotic resistance, carbon source utilization had false positive rates ranging from 20% to 90% (Griffith et al., 2003; Harwood et al., 2003; Myoda et al., 2003).

The accuracy and reproducibility of the *E. coli mdh* assay was further confirmed by the results of a blind study of water samples spiked with in-library *E. coli* isolates from dog, horse and seagull ($n=9$) which had host specific target sequences. The blind sequencing and analysis of these library isolates generated *mdh* target sequences that were 100% accurate and 100% reproducible in both sequence identity and matching blinded isolates to the correct host. No other library-based MST method approaches the levels of accuracy and reproducibility for blinded replicates found for the *mdh* assay. For example, reproducibility, as assessed by analysis of blinded in-library replicates by two phenotypic and five genotypic MST methods, ranged from 13% to 100% for an eight-way host comparison and 0% to 100% for a human versus non-human comparison (Stoeckel et al., 2004a,b). Only pulsed field gel electrophoresis (PFGE) was 100% accurate in identifying replicates, while antibiotic resistance analysis, ribotyping, BOX-PCR, REP-PCR and carbon utilization profiling failed to identify replicates with high accuracy and reproducibility.

The results of blinded analyses of non-library isolates and of geographical variation confirmed the validity and accuracy of the *mdh* library. It has been reported that analysis of non-library isolates provides a more realistic assessment of the accuracy of a library-based MST method than does analysis of in-library isolates (Moore et al., 2005). The *mdh* assay outperforms antibiotic resistance analysis and ribotyping MST assays (Moore et al., 2005) in this regard. The former assay had positive predictivities and sensitivities ranging from 58% to 92% (Table 2), while antibiotic resistance analysis and ribotyping had rates of correct prediction for non-library isolates ranging from 6% to 67% (Moore et al., 2005). The sequences from 95% (18/19) of the blinded non-library isolates from deer, horse, seagull and dog collected in California each exactly matched one or more in-library sequences for the corresponding host. The degree of geographical variation of the *mdh* assay

appeared to be relatively low; i.e. 91% of *E. coli* isolates from horse, dog, and deer fecal samples collected in Florida matched one or more library sequences for fecal samples from the corresponding host collected in California. The low degree of geographical variation of the *mdh* assay may reflect several factors: First, Mdh is an essential metabolic enzyme, and a short 150 bp sequence of the highly conserved catalytic domain is analyzed. Second, there appears to be a low degree of variability among *mdh* sequences from isolates from a given fecal sample; only one to two different *mdh* sequences were found in up to five isolates per fecal sample from hosts such as human, dog and horse, where each fecal sample was collected from a single individual. Finally, the accuracy and reproducibility of sequencing the *mdh* gene target was found to be 100%, which greatly exceeds the reproducibility of other MST methods, such as ribotyping and antibiotic resistance analysis for blinded replicates.

The low geographical diversity of the *mdh* gene target and high degrees of accuracy and reproducibility of the assay suggest that (i) the library size for this assay could be considerably smaller than required for other MST methods, (ii) the *mdh* library generated for this study, although relatively small in size (ca. 300 isolates) may be large enough to be representative, and (iii) therefore the parameters for host identification given in Table 2 and summarized above may be applicable to non-library isolates. Library development for MST is a tedious, expensive exercise, and it has been proposed that for MST assays such as antibiotic resistance analysis and DNA ribotyping, representative libraries may need to be extremely large and contain isolates from a broad geographic region (Griffith et al., 2003; Harwood et al., 2003; Scott et al., 2003; Wiggins et al., 2003). An MST assay, such as the *mdh* assay, that exhibits low geographical diversity and can utilize a relatively small library while remaining applicable to geographically diverse locations offers advantages of faster, less expensive and less labor intensive library development.

The applicability of the *mdh* assay to environmental isolates was demonstrated on isolates from a telephone company vault containing standing water characterized by high levels of fecal indicators from unknown host sources. The results of the *mdh* gene assay, namely that four of five isolates matched seagull target sequences, are consistent with other data: First, vault samples showed no measurable human fecal contamination (C_ts > 37.5) in a quantitative PCR assay for the human specific *esp* gene target of *E. faecium* reported by Scott et al. (2005), but had high levels of the *16S* rRNA reference gene (C_t = 15) which is diagnostic for the levels of total *Enterococci*. Positive controls (human sewage samples), showed strikingly greater levels of the human specific target (C_ts = 27), with C_t values and thus levels of the *16S* rRNA reference target comparable to those found in the vault. Thus, at least a 10,000-fold lower concentration of the human specific target was found in vault isolates compared to sewage isolates with comparable levels of total *Enterococci*. Second, the local wastewater agency had performed dye testing and line pressure testing of nearby wastewater lines, which indicated that there was no leakage from their systems (John Simonetti, Westbay Sanitary District, personal communication). Thus, all of the data is consistent with seagull or avian fecal contamination of the vault and appears to exclude the possibility of significant human fecal contamination.

The *mdh* gene target appears to have several advantages for MST compared to the only other DNA sequencing MST method, i.e. analysis of the *E. coli uidA* (*gusA*) gene which codes for β-glucuronidase. First, the *mdh* gene is not subject to horizontal gene transfer (Boyd et al., 1994; Pupo et al., 1997), while no reports on the existence or absence of horizontal gene transfer in the vicinity of the *gus*R, A, B, C operon or on comparison of the phylogenetic trees constructed by *gusA* and MLEE were found. Since the validity of clonal theory is the cornerstone of many MST methods, it is essential to demonstrate that genetic diversity at a given gene locus is representative of *E. coli* evolutionary history. Second, the most populated *uidA* allele, *uidA1*, which accounted for 36% of environmental isolates, had to be eliminated from MST analysis, since it was found in isolates from numerous host groups (Ram et al., 2004). Finally, the frequencies of alleles for the *uidA* gene are not evenly distributed. The predominant alleles *uidA1* and *uidA2* are found in numerous hosts, and the most discriminative alleles, such as *uidA5* and *uidA11* which are found only in birds, represent a small fraction of total isolates and appear to be relatively uncommon in environmental isolates.

By comparison, judging from our dendrograms, the various alleles at the *mdh* locus of *E. coli* are more evenly distributed, such that we did not observe any predominant allele which is frequently found in all host species and accounts for a significant portion of environmental isolates. In addition, we did not cull particular alleles for our MST assay; by treating each *mdh* allele as equally informative, the sensitivity, specificity and practicality of our assay can be more objectively assessed.

The relative degrees of sequence homogeneity by host for the *mdh* and *uidA* genes were similar, with high

levels of sequence homogeneity in isolates from horse, followed by dog, and lower levels of homogeneity in human and seagull (see Results) (Ram et al., 2004). In contrast, ribotyping and antibiotic resistance analysis found high levels of homogeneity in humans and dogs and markedly lower levels of homogeneity in horse (Anderson, 2003; Stoeckel et al., 2004a,b), indicating that the individual target and/or assay method may impact the observed level of host diversity.

In addition to investigating the diversity of the *mdh* gene target and its suitability for MST, the polymorphisms of the Mdh protein were examined to ascertain if there were structural features that rendered the NAD+ binding domain more amenable to mutation than the catalytic domain. The structure of the NAD+ binding domain is highly conserved across members of the dehydrogenase enzyme family, which includes enzymes with diverse substrate specificities, in a range of organisms. For example, Mdh and lactate dehydrogenase (Ldh), which are both type A dehydrogenases and possess structurally homologous NAD+ binding and catalytic domains, share only 20% sequence identity (Hall et al., 1992).

The NAD+ binding domain is able to retain its three-dimensional structure and enzymatic activity despite a large number of mutations, due to the manner by which it interacts with NAD+. When structures of malate dehydrogenase bound to a substrate analog (citrate) in the presence and absence of NAD+ (1EMD and 2CMD, respectively) are structurally aligned, the main chain atoms of both structures are essentially superimposable (RMSD 0.112 Å). Because the majority of the hydrogen bonds to the NAD+ cofactor are derived from stationary backbone amide and carbonyl groups, the minor side chain conformational changes accompanying the binding of NAD+ to MDH do not affect the protein–ligand interactions. Extensive hydrogen bonding interactions with backbone protein atoms results in less reliance on amino acid side chains to confer cofactor binding specificity, and in heterogeneous amino acid sequences among members of protein families (Hall and Banaszak, 1993).

In addition to primarily being localized to the NAD+ binding domain of Mdh, all of the polymorphic sites were located in surface exposed regions of the protein, which are characterized by increased main chain and side chain temperature factors relative to the remainder of the protein. Surface exposed residues tend to be the most variable regions of proteins due to a lack of conformational restriction on the amino acid side chains. The interior of the protein requires precise packing of amino acid side chains in order to assume a correct fold, whereas amino acids on the exterior are allowed increased conformational freedom, so long as protein structure and function are not perturbed (Creighton, 1993). The high thermal factors of regions containing polymorphic residues may reflect that flexible regions in the protein may be more mutable than the less flexible segments.

Since Ldh and Mdh are members of the same subclass of dehydrogenases and have structurally homologous NAD+ and catalytic domains (Hall et al., 1992), and *mdh* shows a degree of success as a target for identifying specific host sources of fecal pollution by MST, it is possible that structural homologs in the dehydrogenase family, such as Ldh, may also be suitable targets for MST, provided that they meet the criteria for sub-typing of microbial organisms by DNA sequencing, e.g. possess low rates of horizontal gene transfer and recombination, and have single gene copies in the genome or multiple identical gene copies.

In conclusion, we have demonstrated the validity and efficacy of identifying gene targets by the criteria for sub-typing of microbial organisms by DNA sequencing (Olive and Bean, 1999) and one additional criterion developed by the authors, and the applicability and advantages of DNA sequencing technology to MST. We propose that the *mdh* MST assay would be most effective and accurate in urban watersheds with limited sources of pollution potentially from the hosts studied here, or when applied in conjunction with host specific assays, such as the human specific *E. faecium* target and assay developed by Scott et al. (2005). Further, we support the proposal (Scott et al., 2002) that a toolbox of assays that directly sample gene sequences of different targets should be developed in order to facilitate MST of multiple hosts with high levels of accuracy and reproducibility. With local knowledge of potential sources of fecal pollution, it would be possible to make an informed choice of which methods from the toolkit would best suit an individual watershed.

Acknowledgements

Sophie Archambeault, Katherine Clark, Brooke Finkmoore and Jennifer O'Shaughnessy for technical assistance, Carolann Towe for collection of fecal samples from seagull flocks, Ranger Douglas Heisinger of the San Pedro Park and Matthew Woodworth and Robert Gonzalez of the University of California San Francisco Biomolecular Resource Center for collection of fecal samples, Douglas Coffman for isolation of *E. coli* colonies, and Bernard Halloran for initiating the project that derived funding for this study. Molecular

graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR-01081).

Funding for this project has been provided in full or in part through an Agreement with the State Water Resources Control Board (SWRCB) pursuant to the Costa-Machado Water Act of 2000 (Proposition 13) and any amendments thereto for the implementation of California's Nonpoint Source Pollution Control Program. The contents of this document do not necessarily reflect the views and policies of the SWRCB, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

References

Anderson, M.A., 2003. Frequency distributions of *Escherichia coli* subtypes in various fecal sources over time and geographical space: application to bacterial source tracking methods. Department of Biology. Tampa, FL, University of South Florida. M. S., 109.

Applied Biosystems, 2000. Automated DNA Sequencing. Foster City, CA.

Applied Biosystems, 2001. ABI PRISM® BigDye™ Terminator v3.0 Ready Reaction Cycle Sequencing Kit protocol. Foster City, CA.

Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., Shao, Y., 1997. The complete genome sequence of *Escherichia coli* K-12. Science 277, 1453–1474.

Boyd, E.F., Nelson, K., Wang, F.S., Whittam, T.S., Selander, R.K., 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (mdh) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. U. S. A. 91, 1280–1284.

Cabelli, V.J., 1977. Indicators of recreational water quality. In: Hoadley, A.W., Dutka, B.J. (Eds.), Bacterial Indicators/Health Hazards Associated with Water. American Society for Testing and Materials. Philadelphia. ASTM STP, vol. 635, pp. 222–238.

Cabelli, V.J., 1983. Health effects for marine recreation waters. USEPA 600/1-80-031. Research Triangle Park, NC, Health Effects Research Laboratory.

Cabelli, V.J., Dufour, A.P., McCabe, L.J., Levin, M.A., 1982. Swimming-associated gastroenteritis and water quality. Am. J. Epidemiol. 115, 606–616.

Carson, C.A., Shear, B.L., Ellersieck, M.R., Schnell, J.D., 2003. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. Appl. Environ. Microbiol. 69, 1836–1839.

Creighton, T.E., 1993. Proteins: Structures and Molecular Properties, 2nd ed. W.H. Freeman and Company, New York.

Don, R.H., Cox, P.T., Wainwright, B.J., Baker, K., Mattick, J.S., 1991. 'Touchdown' PCR to circumvent spurious priming during gene amplification. Nucleic Acids Res. 19, 4008.

Dufour, A.P., 1984. Health effects criteria for fresh recreational waters. U. Office of Research and Development. EPA-600/1-84-004.

Dufour, A.P., Ballentine, R., 1986. Ambient water quality criteria for bacteria. EPA 440-5-84-002. Washington, D.C., Office of Research and Development, U.S. Environmental Protection Agency.

Escobar-Páramo, P., Sabbagh, A., Darlu, P., Pradillon, O., Vaury, C., Denamur, E., Lecointre, G., 2004. Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study. Mol. Phylogenet. Evol. 30, 243–250.

Geldreich, E.E., 1966. Sanitary significance of fecal coliform in the Environment. Water Control Research Series Publication WP-20-3. U. S. D. o. t. Interior. Cincinnati, Ohio, Federal Water Pollution Control Administration.

Gordon, D.M., Lee, J., 1999. The genetic structure of enteric bacteria from Australian mammals. Microbiology 145, 2673–2682.

Griffith, J.F., Weisberg, S.B., McGee, C.D., 2003. Evaluation of microbial source tracking methods using mixed fecal sources in aqueous test samples. J. Water Health 1, 141–151.

Guan, S., Xu, R., Chen, S., Odumeru, J., Gyles, C., 2002. Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources. Appl. Environ. Microbiol. 68, 2690–2698.

Hall, M.D., Banaszak, L.J., 1993. Crystal structure of a ternary complex of *Escherichia coli* malate dehydrogenase citrate and NAD at 1.9 A resolution. J. Mol. Biol. 232, 213–222.

Hall, M.D., Levitt, D.G., Banaszak, L.J., 1992. Crystal structure of *Escherichia coli* malate dehydrogenase. A complex of the apoenzyme and citrate at 1.87 A resolution. J. Mol. Biol. 226, 867–882.

Harwood, V.J., Wiggins, B., Hagedorn, C., Ellender, R.D., Gooch, J., Kern, J., Samadpour, M., Chapman, A.C., Robinson, B.J., Thompson, B.C., 2003. Phenotypic library-based microbial source tracking methods: efficacy in the California collaborative study. J. Water Health 1, 153–166.

Higgins, D.G., Thompson, J.D., Gibson, T.J., 1996. Using CLUSTAL for multiple sequence alignments. Methods Enzymol. 266, 383–402.

Ivanetich, K.M., Reid, R.C., Ellison, R., Perry, K., Taylor, R., Reschenberg, M., Mainieri, A., Zhu, D., Argo, J., Cass, D., Strickland, C., 1999. Automated purification and quantification of oligonucleotides. Biotechniques 27, 810–812, 814–818, 820 passim.

Lecointre, G., Rachdi, L., Darlu, P., Denamur, E., 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. Mol. Biol. Evol. 15, 1685–1695.

Meays, C.L., Broersma, K., Nordin, R., Mazumder, A., 2004. Source tracking fecal bacteria in water: a critical review of current methods. J. Environ. Manag. 73, 71–79.

Moore, D.F., Harwood, V.J., Ferguson, D.M., Lukasik, J., Hannah, P., Getrich, M., Brownell, M., 2005. Evaluation of antibiotic resistance analysis and ribotyping for identification of faecal pollution sources in an urban watershed. J. Appl. Microbiol. 99, 618–628.

Myoda, S.P., Carson, C.A., Fuhrmann, J.J., Hahm, B.K., Hartel, P.G., Yampara-Lquise, H., Johnson, L., Kuntz, R.L., Nakatsu, C.H., Sadowsky, M.J., Samadpour, M., 2003. Comparison of genotypic-based microbial source tracking methods requiring a host origin database. J. Water Health 1, 167–180.

Nelson, K., Selander, R.K., 1992. Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. 174, 6886–6895.

Nelson, K., Whittam, T.S., Selander, R.K., 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (gapA) in natural populations of Salmonella

and *Escherichia coli*. Proc. Natl. Acad. Sci. U. S. A. 88, 6667–6671.

Noble, R.T., Moore, D.F., Leecaster, M.K., McGee, C.D., Weisberg, S. B., 2003. Comparison of total coliform, fecal coliform, and enterococcus bacterial indicator response for ocean recreational water quality testing. Water Res. 37, 1637–1643.

NRDC, 2004. Testing the Waters 2004: A Guide to Water Quality at Vacation Beaches.

Ochman, H., Selander, R.K., 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. 157, 690–693.

O'Shaughnessy, J.B., Chan, M., Clark, K., Ivanetich, K.M., 2003. Primer design for automated DNA sequencing in a core facility. Biotechniques 35, 112–116, 118–121.

Olive, D.M., Bean, P., 1999. Principles and applications of methods for DNA-based typing of microbial organisms. J. Clin. Microbiol. 37, 1661–1669.

Orskov, F., Orskov, I., 1981. Enterobacteriaceae. In: Broude, A.I. (Ed.), Medical Microbiology and Infectious Diseases. W.B. Saunders Co, Philadelphia, pp. 340–352.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. J. Comput. Chem. 25, 1605–1612.

Poirot, O., Suhre, K., Abergel, C., O'Toole, E., Notredame, C., 2004. 3DCoffee: a web server for mixing Sequences and Structures into multiple sequence alignments. Nucleic Acids Res. 32, W37–W40 (Web Server issue).

Pruss, A., 1998. Review of epidemiological studies on health effects from exposure to recreational water. Int. J. Epidemiol. 27, 1–9.

Pupo, G.M., Karaolis, D.K., Lan, R., Reeves, P.R., 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and mdh sequence studies. Infect. Immun. 65, 2685–2692.

Ram, J.L., Ritchie, R.P., Fang, J., Gonzales, F.S., Selegean, J.P., 2004. Sequence-based source tracking of *Escherichia coli* based on genetic diversity of beta-glucuronidase. J. Environ. Qual. 33, 1024–1032.

Scott, T.M., Rose, J.B., Jenkins, T.M., Farrah, S.R., Lukasik, J., 2002. Microbial source tracking: current methodology and future directions. Appl. Environ. Microbiol. 68, 5796–5803.

Scott, T.M., Parveen, S., Portier, K.M., Rose, J.B., Tamplin, M.L., Farrah, S.R., Koo, A., Lukasik, J., 2003. Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. 69, 1089–1092.

Scott, T.M., Jenkins, T.M., Lukasik, J., Rose, J.B., 2005. Potential use of a host associated molecular marker in *Enterococcus faecium* as an index of human fecal pollution. Environ. Sci. Technol. 39, 283–287.

Simpson, J.M., Santo Domingo, J.W., Reasoner, D.J., 2002. Microbial source tracking: state of the science. Environ. Sci. Technol. 36, 5279–5288.

Sinton, L.W., Finlay, R.K., Hannah, D.J., 1998. Distinguishing human from animal faecal contamination in water: a review. N. Z. J. Mar. Freshw. Res. 32, 323–348.

Stoeckel, D.M., Kephart, C.M., Harwood, V.J., Anderson, M.A., Dontchev, M., 2004a. Diversity of fecal indicator bacteria subtypes: implications for construction of microbial source tracking libraries. American Society for Microbiology General Meeting, New Orleans, LA.

Stoeckel, D.M., Mathes, M.V., Hyer, K.E., Hagedorn, C., Kator, H., Lukasik, J., O'Brien, T.L., Fenger, T.W., Samadpour, M., Strickler, K.M., Wiggins, B.A., 2004b. Comparison of seven protocols to identify fecal contamination sources using *Escherichia coli*. Environ. Sci. Technol. 38, 6109–6117.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

USEPA, 1986. Ambient water quality criteria for bacteria. EPA-440-5-84-002. USEPA.

USEPA, 2003. Bacterial Water Quality Standards for Recreational Waters (Freshwater and Marine Waters) EPA-823-R-03-008. U. S. E. P. A. Office of Water. Washington, D.C., Office of Water, U.S. Environmental Protection Agency.

Wang, F.S., Whittam, T.S., Selander, R.K., 1997a. Evolutionary genetics of the isocitrate dehydrogenase gene (*icd*) in *Escherichia coli* and *Salmonella enterica*. J. Bacteriol. 179, 6551–6559.

Wang, H., Yang, C.H., Lee, G., Chang, F., Wilson, H., del Campillo-Campbell, A., Campbell, A., 1997b. Integration specificities of two lambdoid phages (21 and e14) that insert at the same attB site. J. Bacteriol. 179, 5705–5711.

Wiggins, B.A., Cash, P.W., Creamer, W.S., Dart, S.E., Garcia, P.P., Gerecke, T.M., Han, J., Henry, B.L., Hoover, K.B., Johnson, E.L., Jones, K.C., McCarthy, J.G., McDonough, J.A., Mercer, S.A., Noto, M.J., Park, H., Phillips, M.S., Purner, S.M., Smith, B.M., Stevens, E.N., Varner, A.K., 2003. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. Appl. Environ. Microbiol. 69, 3399–3405.

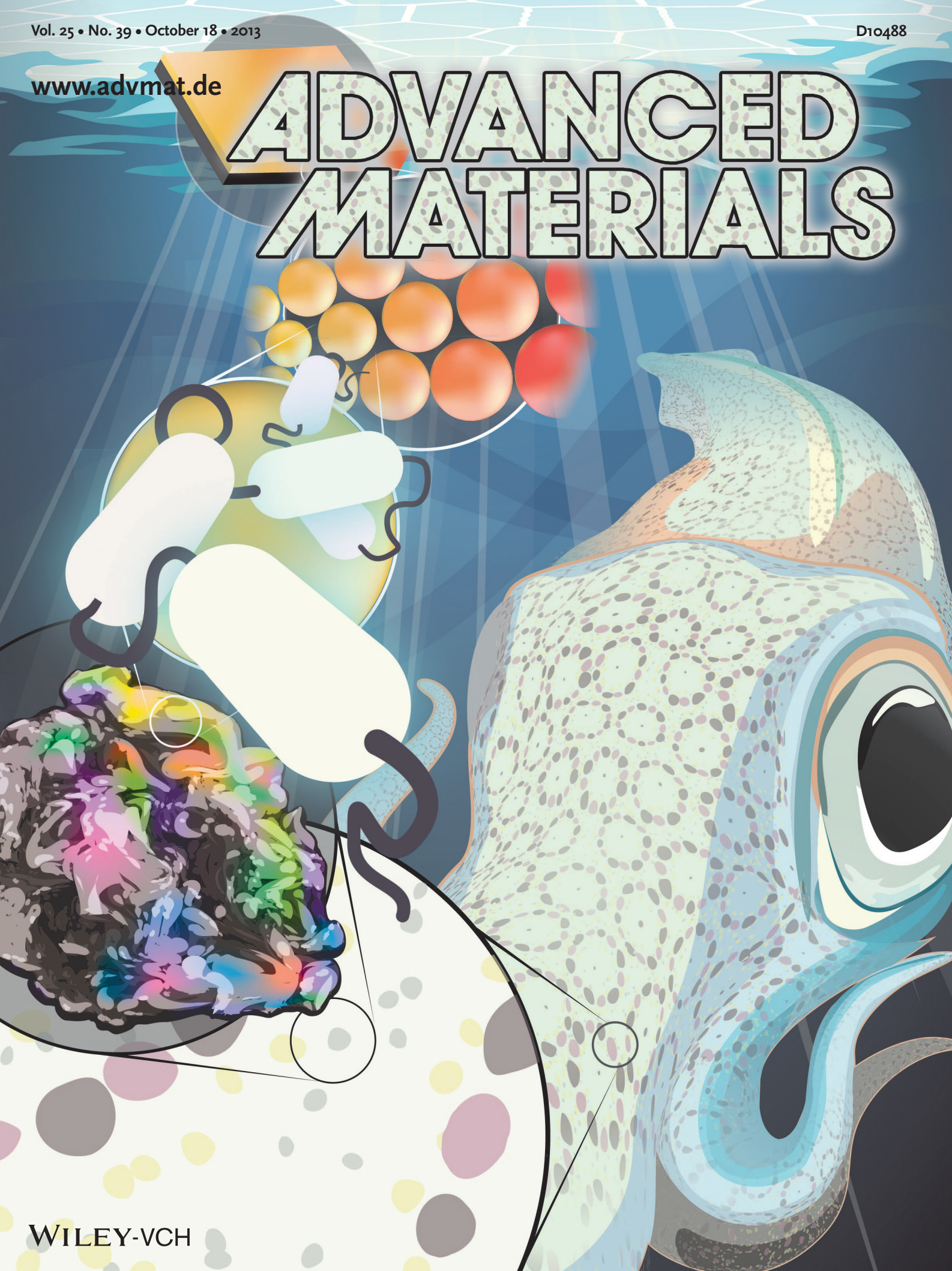
Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. U. S. A. 87, 4576–4579.

Vol. 25 • No. 39 • October 18 • 2013

D10488

www.advmat.de

ADVANCED MATERIALS



WILEY-VCH

Reconfigurable Infrared Camouflage Coatings from a Cephalopod Protein

*Long Phan, Ward G. Walkup IV, David D. Ordinario, Emil Karshalev, Jonah-Micah Jocson, Anthony M. Burke, and Alon A. Gorodetsky**

Over millions of years of evolution, nature has mastered the manipulation of light and color through the controlled arrangement of nanoscale structures.[1–3] Structural color is widespread in the animal kingdom (noteworthy examples are present in insects, reptiles, cephalopods, frogs, fish, and birds), and it typically plays a crucial biological role in camouflage, communication, and temperature control.[1–3] The functional diversity of structural color has inspired the development of materials and surfaces that imitate natural coloration mechanisms.[4–6] Such studies have shown promise for applications in commercial and military technologies, including reconfigurable camouflage, advanced adaptable optics, and energy efficient devices.[4–11]

For stealth applications,[7–9] the camouflage mechanisms of cephalopods are particularly attractive. For example, squid in the Loliginidae family possess specialized dermal cells known as iridocytes.[11–19] These cells contain alternating layers of cell membrane-enclosed platelets composed of a structural protein called reflectin, which has a high refractive index, and extracellular space, which has a low refractive index.[11–19] The lamellar structures function as modular Bragg reflectors, whose spacing Loliginids tune to modulate their skin coloration across the entire visible spectrum and even into the near infrared.[11–19] Thus, the Loliginidae family demonstrates two capabilities that are highly desirable in next generation infrared stealth technology:[7–9] the reflectance of light over a broad range of wavelengths (extending from the visible region into the near infrared region) and reversible dynamic color tunability.

Herein, we describe our progress towards Loliginid-inspired biomimetic infrared camouflage. Our strategy uses simple fabrication techniques to prepare thin films from a water-soluble *Loligo (Doryteuthis) pealeii* reflectin isoform[15] on graphene oxide-coated substrates. We can dynamically tune the reflectance of our coatings over more than 600 nm, which, to the best of our knowledge, represents the largest tunable reflectance window for any biological thin film. This tunability allows our films to reversibly disappear and reappear when visualized with an infrared imaging camera. Our findings represent an inexpensive and robust strategy for the fabrication of dynamically tunable infrared camouflage coatings on arbitrary surfaces.

L. Phan, D. D. Ordinario, E. Karshalev, J.-M. Jocson, A. M. Burke, Prof. A. A. Gorodetsky
Department of Chemical Engineering and Materials Science
University of California
Irvine, Irvine, CA 92697, (USA)
E-mail: alon.gorodetsky@uci.edu
W. G. Walkup IV
Division of Biology
California Institute of Technology
Pasadena, CA 91125, (USA)

DOI: 10.1002/adma.201301472

We began our studies by optimizing heterologous expression and inclusion body preparation protocols for the production of the histidine-tagged reflectin A1 (RfA1) isoform in *E. coli*. The protein was sequentially purified by immobilized metal affinity chromatography (IMAC) under denaturing conditions and high performance liquid chromatography (HPLC) (Figure S1 and Figure S2, Supporting Information). The identity of the purified RfA1 was definitively confirmed by in-gel tryptic digestion and tandem mass spectrometry, which produced sequence coverage in excess of 96% (Figure S3). Notably, our optimized expression and purification procedure yielded >800 mg of purified RfA1 per liter of *E. coli* cell culture, with a purity of over 99%, as judged from HPLC chromatograms (Figure S1). The high yield and excellent purity facilitated subsequent thin film fabrication.

We proceeded to screen a variety of substrates and surface treatments for the reliable formation of RfA1 thin films. The best results were obtained by spincasting 5 to 10 nm films of graphene oxide on (3-aminopropyl)triethoxy silane (APTES)-modified fused silica (or glass) substrates (**Figure 1**A). This material, which presents a negatively charged, amphipathic surface, is well suited for the electrostatically-induced assembly of positively charged RfA1.[20,21] We used a modified doctor blading approach to spread RfA1 onto the graphene oxide-coated substrates (Figure 1A), yielding smooth films over centimeter areas, as determined by optical microscopy and atomic force microscopy (AFM) (Figure 1B and Figure S4).

We next investigated the optical properties of our RfA1 films. Here, the distinct coloration of the surfaces was dictated by the film thickness. As examples, Figure 1B and Figure 1C show typical optical images and reflectance spectra for two RfA1 films on silica: a 125 nm-thick blue film and a 207 nm-thick orange film, as determined by profilometry and interferometry. The observed peak reflectance wavelengths (λ = 431 nm and λ = 625 nm for the blue and orange films, respectively) yielded theoretical thicknesses (138 nm and 204 nm for the blue and orange films, respectively) that were in excellent agreement with the measured values (see Supporting Information for calculation details).[22]

Inspired by the dynamic optical properties of reflectin nanostructures, we sought to shift the reflectance of our RfA1 films into the infrared region of the electromagnetic spectrum. Given that Loliginids can dynamically modulate their skin reflectance across the entire visible spectrum and even out to near infrared wavelengths of ≈800 nm,[19] we postulated that it should also be

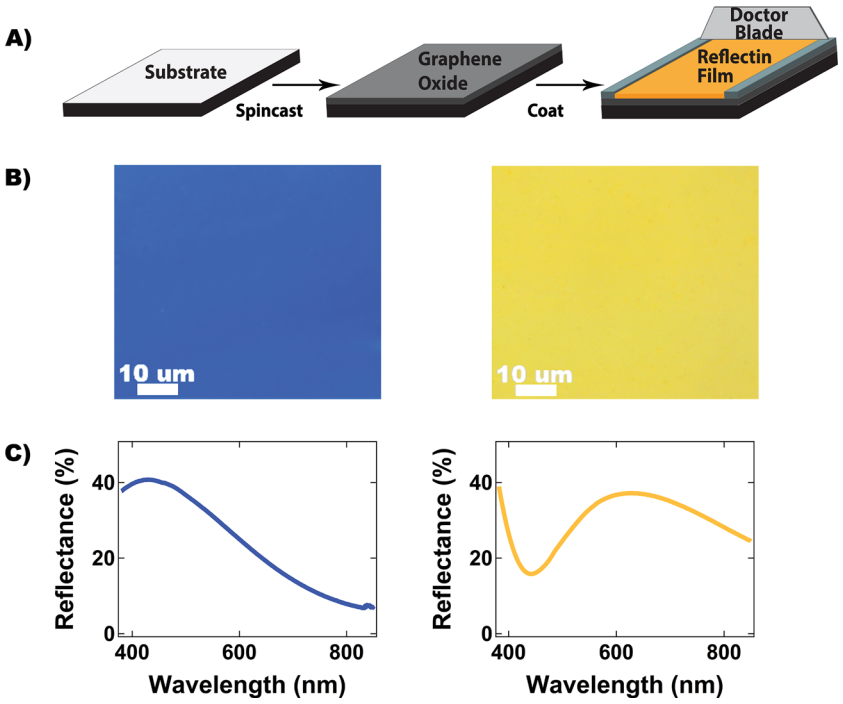


Figure 1. A) Illustration of the fabrication of RfA1-coated substrates. The process consists of casting a graphene oxide film onto an APTES-modified substrate, followed by doctor blading of RfA1 directly onto this surface. B) Typical optical microscopy images for blue (left) and orange (right) RfA1 thin films, with thicknesses of 125 nm and 207 nm, respectively. C) Correponding reflectance spectra for the blue (left) and orange (right) films.

possible to tune the reflectance of our RfA1 thin films across a similar, or even larger, wavelength range. Thus, we sought conditions that would significantly increase the thickness of our RfA1 films and, consequently, shift their reflectance spectra toward the infrared.

We first explored the response of our coatings to an increase in relative humidity. Previous work demonstrated that reflectin thin films exhibit hydrogel-like behavior and can be swollen by exposure to water vapor, modulating the peak reflectance across the visible spectrum.[14,17] Indeed, for a typical orange film, the presence of water vapor (relative humidity of 75%) caused the thickness to increase from 207 nm to 244 nm, with a concomitant peak reflectance shift from $\lambda = 625$ nm to $\lambda = 765$ nm, as determined by interferometry (Figure S5). These observations were in qualitative agreement with previous studies.[17]

To modulate the reflectance of our RfA1 coatings over an even larger wavelength range, we explored their response to a variety of chemical stimuli (see Supporting Information). We discovered that exposing our films to vapor from a concentrated acetic acid solution induced a large, reversible shift in the reflectance spectra (**Figure 2**A and Figure S6). For example, when exposed to acetic acid vapor, an orange film on glass exhibited a new peak centered at $\lambda = 1200$ nm (Figure 2A), which correlated to a film thickness of 394 nm (see Supporting Information for calculation details). Bovine serum albumin (BSA) control films prepared and treated under identical conditions displayed a much smaller response to acetic acid vapor (Figure S7). Surprisingly, the RfA1 films were stable upon exposure to acetic acid and could be dynamically cycled between visible and near infrared reflectance more than 140 times prior to degradation (Figure S4 and Figure S6). The large wavelength range and exceptional stability of our material represent important advantages for infrared camouflage coating applications.[7–9]

To explain the observed acid-induced swelling behavior, we considered the structure of the RfA1 films, which is illustrated in Figure 2B. The films are known to consist of closely packed RfA1 nanoparticles, as demonstrated by TEM and SAXS studies.[15] Thus, we investigated the formation and size of RfA1 nanoparticles in solution as a function of pH via circular dichroism (CD), fluorescence spectroscopy, and dynamic light scattering (DLS). In deionized water at a pH of ≈6.3, DLS revealed the formation of a stable solution of aggregated RfA1 with a narrowly distributed hydrodynamic radius of $R_H = 75$ nm (±11 nm) (Figure S8). Furthermore, CD indicated that the aggregated RfA1 possessed little secondary structure (Figure S9) and fluorescence spectroscopy revealed a peak at $\lambda = 342$ nm, which is characteristic of tryptophan residues (Figure S10).[23] Decreasing the pH to ≈2.3 caused little change in the protein structure, as evidenced by the CD spectra (Figure S9) and fluorescence measurements (Figure S10).

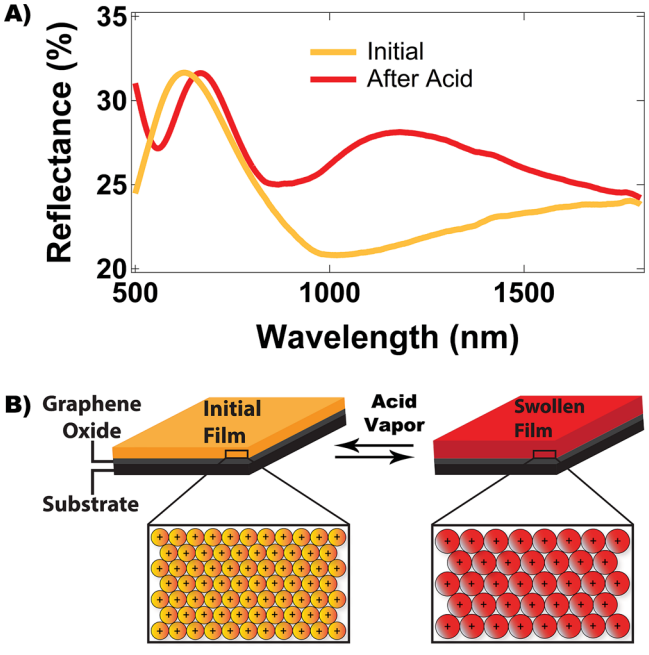


Figure 2. A) Effect of acetic acid vapor on the reflectance spectrum of an RfA1 film. Swelling of the film leads to the appearance of a large reflectance peak centered at 1200 nm. The curves have been smoothed for clarity. B) Illustration of acetic acid vapor-induced swelling for a film composed of net positively charged RfA1 nanoparticles.

However, DLS revealed that the average size of the aggregates increased to R_H = 99 nm (±15 nm) at low pH, with a narrow distribution and no evidence of nanoparticle break-up (Figure S8). In their totality, these observations indicated that RfA1 nanoparticles occupy a larger volume under more acidic conditions, though the protein itself undergoes minimal structural changes.

Given previous literature precedent[15] and our own observations, we postulate a simple model for acid-induced swelling of our films (Figure 2B). The isoelectric point (pI) of RfA1 is 9.1, so at neutral pH, RfA1 possesses a net positive charge.[15] Acidic conditions further increase the positive charge of the RfA1 protein, likely through the protonation and neutralization of the side chains of aspartic and glutamic acid residues (pKa of ≈3.9 and ≈4.3, respectively) and through the protonation of the imidazole groups of histidine residues (pKa of ≈6.0 to ≈7.0).[23] This increase in net positive charge leads to enhanced intra-protein electrostatic repulsion and accounts for the observed swelling of the RfA1 nanoparticles in solution. Since our films consist of closely packed RfA1 nanoparticles,[15] the swelling of the individual nanoparticles explains the significant increase in film thickness under acidic conditions.

In a proof of principle demonstration of our coatings' infrared camouflage capabilities, we visualized RfA1-covered silica substrates with a camera in both standard and infrared imaging modes, while modulating their reflectance in situ (**Figure 3**A). Under white light illumination, a typical RfA1 film appeared orange and could be readily distinguished from a piece of black paper and a leaf (Figure 3B). Under infrared illumination at λ = 940 nm without any chemical stimulus, the RfA1 coating appeared black, with a relative brightness of 13% and resembled the paper control sample, with a relative brightness of 1% (Figure 3B and Figure S11). The leaf, which reflected in the infrared as expected for typical foliage, appeared red, with a relative brightness of 57% (Figure 3B and Figure S11). However, under infrared illumination at λ = 940 nm in the presence of acetic acid vapor, the reflectance of the RfA1 film shifted into the near infrared, and the RfA1-coated substrate appeared red, with a maximum relative brightness of 100% (Figure 3B and Figure S11). The RfA1-coated substrate thus became distinct from the paper control, with a relative brightness of 7%, and now resembled the leaf control, with a relative brightness of 64%. (Figure 3B and Figure S11). Overall, this experiment demonstrated that the RfA1-covered substrates could be dynamically hidden or revealed when probed with standard infrared detection equipment.

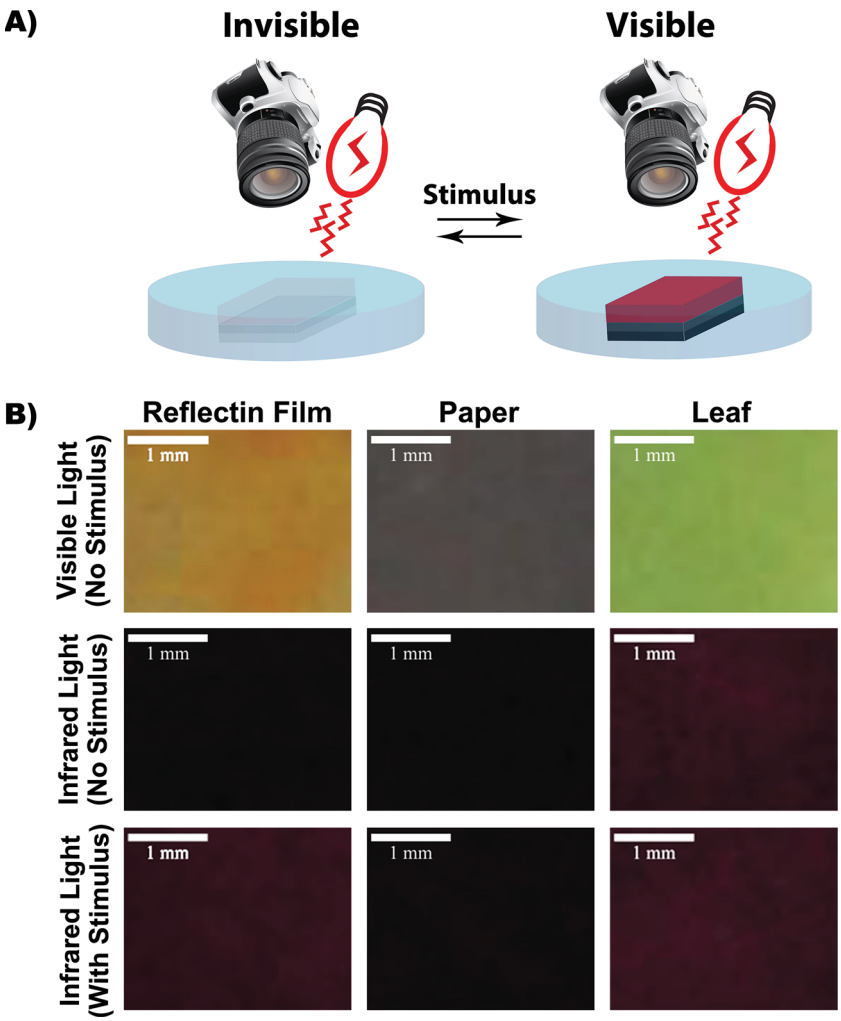


Figure 3. A) Illustration depicting the appearance of the RfA1 film in the absence and presence of an external stimulus (acetic acid), when visualized with an infrared camera. B) First row: images of an RfA1-coated substrate (left), a black piece of paper (center), and a plant leaf (right) that were captured with a standard camera under white light illumination. Second row: images of the same RfA1-coated substrate (left), black piece of paper (center), and plant leaf (right) that were captured with an infrared camera in the absence of acetic acid under infrared diode illumination. Third row: images of the same RfA1-coated substrate (left), black piece of paper (center), and plant leaf (right) that were captured with an infrared camera in the presence of acetic acid under infrared diode illumination. Note the change in apparent coloration from black to red for the RfA1-coated substrate in the absence (second row, left) and presence (third row, left) of acetic acid vapor.

In conclusion, we have demonstrated graphene-templated, biomimetic camouflage coatings that possess several important advantages. The thin film fabrication strategy is inexpensive, robust, and compatible with a wide range of substrates of arbitrary form factor. The reflectance spectra of our coatings are modulated in situ, with an accessible coloration range between λ = 400 nm and λ = 1200 nm, as well as a dynamic tunability of over 600 nm. To the best of our knowledge, the demonstrated in situ reflectance window is the largest reported for any biological thin film. In addition, we are aware of only limited examples of conformable coatings from soft materials that display dynamic tunability of their reflectance,[24–31] with only a single example demonstrating in situ dynamic reflectance modulation

over a larger portion of the electromagnetic spectrum.[24] Furthermore, the thin film reflectance of our samples extends into the important $\lambda = 700$ nm to $\lambda = 1200$ nm spectral region, which matches the standard imaging range of infrared visualization equipment.[7–9] Given these advantages, our dynamically tunable, infrared-reflective films represent a crucial first step towards the development of reconfigurable and disposable biomimetic camouflage technologies for military stealth applications.[7–9]

Experimental Section

Reflectin A1 Expression and Purification: An *E. coli* codon optimized gene coding for 6X histidine-tagged reflectin A1 protein (RfA1) from *Doryteuthis pealeii* (Genbank: ACZ57764.1) was synthesized and cloned into the pJExpress414 vector (DNA2.0). Note that the histidine tag, while required for purification, may alter the properties of the wild type protein. The vector was transformed into BL21(DE3) cells (Novagen). RfA1 was expressed at 37 °C using Overnight Express Instant Terrific Broth (TB) media (Novagen) supplemented with 100 μg mL^{-1} Carbenicillin. RfA1 was completely insoluble when expressed at 37 °C and was sequestered in inclusion bodies prepared using BugBuster (Novagen) according to the manufacturer's suggested protocols. The RfA1 inclusion bodies were then solubilized in denaturing buffer (pH 7.4, 50 mM sodium phosphate, 300 mM sodium chloride, 6 M guanidine hydrochloride) and purified under denaturing conditions on a HisPur Cobalt Resin gravity column (Thermo Scientific) according to the manufacturer's protocols (elution was performed using denaturing buffer supplemented with 250 mM imidazole). The fractions containing RfA1 protein were pooled and concentrated on an Amicon Concentrator (Millipore) before being purified with high performance liquid chromatography (HPLC) on an Agilent 1260 Infinity system using a reverse phase C18 column. The gradient was evolved from 95% Buffer A:5% Buffer B to 5% Buffer A:95% Buffer B at a flow rate of 0.5 mL min^{-1} over 20 minutes (Buffer A: 99.9% water, 0.1% trifluoroacetic acid; Buffer B: 95% acetonitrile, 4.9% water, 0.1% trifluoroacetic acid). The pure RfA1 protein was pooled, flash frozen in liquid nitrogen, and lyophilized. Protein concentrations and yields were quantified via the Bradford protein assay (BioRad) with BSA as the standard.

Reflectin A1 Analysis: Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was performed to assay protein purity and analyze RfA1 yields throughout the purification process. RfA1 samples were analyzed by SDS-PAGE and Gel Code Blue Reagent staining (Thermo Scientific) on an XCell SureLock Mini-cell Electrophoresis System (Invitrogen) using NuPAGE Novex 4–12% Bis-Tris gels (Invitrogen), with either NuPAGE MOPS or MES (Invitrogen) as the running buffer under reducing conditions (Figure S2). Stained protein bands were subjected to in-gel tryptic digestion, which was performed according to literature procedures.[32] After digestion, the RfA1 peptides were separated on C18 resin beads and analyzed by mass spectrometry on a Waters Synapt G2 instrument outfitted with an electrospray ionization source.

Spectroscopic Characterization: RfA1 was analyzed with dynamic light scattering, fluorescence spectroscopy, UV-visible spectroscopy, and circular dichroism. Dynamic light scattering measurements were performed with a Malvern Zeta-Sizer Nano S instrument. Fluorescence spectra were obtained on a Hitachi F4500 Fluorescence Spectrophotometer. UV-visible spectra were obtained on an Cary 50 Spectrophotometer. Circular dichroism spectra were obtained on a Jasco J-810 Spectropolarimeter. All measurements were performed in deionized water (MilliQ). When necessary, the pH was adjusted with acetic acid and measured both in the presence and absence of RfA1.

Thin Film Fabrication: Square chips cut from either silica substrates (SQI) or glass slides (VWR) were etched in piranha solution (1:3 hydrogen peroxide to sulfuric acid) for 45 minutes. The chips were sequentially rinsed with HPLC-grade acetone and isopropanol, prior to being dried with nitrogen. The substrates were subsequently immersed in a 3% (3-Aminopropyl)triethoxysilane (APTES) in acetone solution for ≈2 (glass) or ≈10 (silica) minutes, rinsed with acetone, and dried with nitrogen. The chips were coated with graphene oxide (Graphene Marketplace) at 3000 RPM on a Laurell WS-400-NNP spincoater. For doctor blading of RfA1, the graphene oxide-coated substrates were modified with teflon tape (McMaster-Carr) on two sides, which acted as a spacing rail and allowed the blade to sit above the substrate surface. The RfA1 solution was then cast onto the substrate in front of the blade, which was translated at a constant speed across the surface to produce thin films. To promote water evaporation and uniform film formation, the coating procedure was performed on a hot plate at 80 °C. Thicker films were produced by varying the RfA1 concentration (typically ≈500 μM for blue films) and/or applying multiple sequential coatings.

Physical Characterization of Thin Films: RfA1 thin films were characterized with atomic force microscopy and profilometry. The surface morphology was examined in situ with an Asylum Research MFP-3D AFM outfitted with an Asylum Research Humidity Sensing Cell. The AFM data was analyzed with the Gwyddion software package. The topological scans were rastered at 0.50 Hz and normalized using polynomial subtraction for improved image quality. Film thicknesses were confirmed with a Bruker Dektak XT profilometer by examining trenches scribed directly into the protein film.

Optical Characterization of Thin Films: RfA1 thin films were characterized with optical microscopy and reflectance measurements. Optical images were obtained on a Carl Zeiss Axio Imager A1M Microscope (outfitted with an Epiplan 20X lens, NA = 0.4) and processed with the manufacturer's AxioVision AC4.5 software. Reflectance spectra were measured in the presence and absence of solvent vapor on three instruments: 1) an Ocean Optics SD2000 Miniature Fiber Optic Spectrometer, outfitted with a tungsten halogen white light source and a R400–7 Reflection/Backscattering Probe (effective wavelength range of $\lambda = 340$ nm to $\lambda = 950$ nm); 2) a Perkin-Elmer Lambda 950 Spectrophotometer, outfitted with deuterium and tungsten halogen white light sources and a 60 mm integrating sphere (effective wavelength range of $\lambda = 190$ nm to $\lambda = 3300$ nm); and 3) a Filmetrics F40-NSR microscope system, outfitted with a tungsten halogen white light source (effective wavelength range of $\lambda = 400$ nm to $\lambda = 800$ nm). For measurements in the presence of water or acetic acid vapor, the samples were housed in custom built closed cells fabricated from either a polystyrene petri dish (Fisher Scientific) or from ABS+ plastic fitted with an N-BK7 glass window (Anchor Optics) (Figure S12). Measurements on the Filmetrics F40-NSR microscope system allowed for confirmation of the film thickness determined by AFM and profilometry. The reflectance data were acquired normal to the substrate surface and were referenced to a SpectraLabs Spectralon Reflectance Standard (unless otherwise noted).

Infrared Imaging of Coated Substrates: The infrared imaging experiments employed a Nikon D-70 Camera equipped with a 35 mm f/1.8 AF-S DX Lens (Nikon) and a 52 mm RM-72 infrared filter (Hoya). The infrared illumination light source consisted of a home-built 26 LED array (XtraLED). The images were captured with an exposure time of 1 second, aperture setting of f/8, and ISO of 200. The integrated relative brightnesses were determined with Adobe Photoshop CS6.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

We are grateful to Beniam Berhane in the UCI Mass Spectrometry Facility for assistance with the tryptic digestion and mass spectrometry

experiments. We thank Vineet Nair in the Law Laboratory and Wytze van der Veer in the UCI Laser Spectroscopy Facility for assistance with the spectroscopic measurements. We also thank the groups of Professor Matthew Law and Professor Albert Yee for help with the AFM characterization. Finally, we thank William Sin at Filmetrics for his assistance with the refractive index measurements. The authors acknowledge Delta Hardware and Industry Co., Ltd. and the University of California, Irvine for financial support.

Received: April 3, 2013
Published online: July 30, 2013

[1] S. Kinoshita, S. Yoshioka, *Structural Colors in Biological Systems: Principles and Applications*, Osaka University Press, Osaka, **2005**.
[2] P. Vukusic, J. R. Sambles, *Nature* **2003**, *424*, 852–855.
[3] P. Ball, *Sci. Am.* **2012**, *306*, 74–79.
[4] A. Saito, *Sci. Technol. Adv. Mater.* **2011**, *12*, 064709/1–064709/13.
[5] Y. Zhao, Z. Xie, H. Gu, C. Zhu, Z. Gu, *Chem. Soc. Rev.* **2012**, *41*, 3297–3317.
[6] H. Fudouzi, *Sci. Technol. Adv. Mater.* **2011**, *12*, 064704/1–064704/7.
[7] S. P. Mahulikar, H. R. Sonawane, G. A. Rao, *Prog. Aerosp. Sci.* **2007**, *43*, 218–245.
[8] V. Rubeziene, G. Minkuviene, J. Baltusnikaite, I. Padleckiene, *Mater. Sci.* **2009**, *15*, 173–177.
[9] S. M. Burkinshaw, G. Hallas, A. D. Towns, *Rev. Prog. Color. Relat. Top.* **1996**, *26*, 47–53.
[10] S. Daehne, U. Resch-Genger, O. S. Wolfbeis, eds, *Near-Infrared Dyes for High Technology Applications*, Vol. *52*, Kluwer Academic Publishers, Dordrecht, **1998**.
[11] E. Kreit, L. M. Mäthger, R. T. Hanlon, P. B. Dennis, R. R. Naik, E. Forsythe, J. Heikenfeld, *J. R. Soc. Interface* **2012**, *10*, 1–13.
[12] L. M. Mäthger, E. J. Denton, N. J. Marshall, R. T. Hanlon, *J. R. Soc. Interface* **2009**, *6*, 149–163.
[13] W. J. Crookes, L. L. Ding, Q. L. Huang, J. R. Kimbell, J. Horwitz, M. J. McFall-Ngai, *Science* **2004**, *303*, 235–238.
[14] R. M. Kramer, W. J. Crookes-Goodson, R. R. Naik, *Nat. Mater.* **2007**, *6*, 533–538.
[15] A. R. Tao, D. G. DeMartini, M. Izumi, A. M. Sweeney, A. L. Holt, D. E. Morse, *Biomaterials* **2010**, *31*, 793–801.
[16] M. Izumi, A. M. Sweeney, D. G. DeMartini, J. C. Weaver, M. L. Powers, A. R. Tao, T. V. Silvas, R. M. Kramer, W. J. Crookes-Goodson, L. M. Mäthger, R. R. Naik, R. T. Hanlon, D. E. Morse, *J. R. Soc. Interface* **2010**, *7*, 549–560.
[17] G. Qin, P. B. Dennis, Y. Zhang, X. Hu, J. E. Bressner, Z. Sun, W. J. Crookes-Goodson, R. R. Naik, F. G. Omenetto, D. L. Kaplan, *J. Polym. Sci. Polym. Phys.* **2013**, *51*, 254–264.
[18] D. G. DeMartini, D. V. Krogstad, D. E. Morse, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2552–2556.
[19] L. M. Mäthger, R. T. Hanlon, *Cell Tissue Res.* **2007**, *329*, 179–186.
[20] D. R. Dreyer, S. Park, C. W. Bielawski, R. S. Ruoff, *Chem. Soc. Rev.* **2010**, *39*, 228–240.
[21] Y. Wang, Z. Li, J. Wang, J. Li, Y. Lin, *Trends Biotechnol.* **2011**, *29*, 205–212.
[22] H. A. MacLeod, *Thin-Film Optical Filters*, 3rd ed., Taylor & Francis, Boca Raton, **2001**.
[23] T. E. Creighton, *Proteins: Structures and Molecular Properties*, 2nd ed., W. H. Freeman & Co., New York, **1993**.
[24] Y. Kang, J. J. Walish, T. Gorishnyy, E. L. Thomas, *Nat. Mater.* **2007**, *6*, 957–960.
[25] C. Kang, E. Kim, H. Baek, K. Hwang, D. Kwak, Y. Kang, E. L. Thomas, *J. Am. Chem. Soc.* **2009**, *131*, 7538–7539.
[26] J. Yoon, W. Lee, E. L. Thomas, *Macromolecules* **2008**, *41*, 4582–4584.
[27] J. J. Walish, Y. Kang, R. A. Mickiewicz, E. L. Thomas, *Adv. Mater.* **2009**, *21*, 3078–3081.
[28] E. Kim, C. Kang, H. Baek, K. Hwang, D. Kwak, E. Lee, Y. Kang, E. L. Thomas, *Adv. Funct. Mater.* **2010**, *20*, 1728–1732.
[29] E. P. Chan, J. J. Walish, E. L. Thomas, C. M. Stafford, *Adv. Mater.* **2011**, *23*, 4702–4706.
[30] Y. Ahn, E. Kim, J. Hyon, C. Kang, Y. Kang, *Adv. Mater.* **2012**, *24*, 127–130.
[31] H. S. Lim, J. H. Lee, J. J. Walish, E. L. Thomas, *ACS Nano.* **2012**, *6*, 8933–8939.
[32] A. Shevchenko, H. Tomas, J. Havlis, J. V. Olsen, M. Mann, *Nat. Protoc.* **2006**, *1*, 2856–2860.

ADVANCED MATERIALS

Supporting Information

for *Adv. Mater.*, DOI: 10.1002/adma.201301472

Reconfigurable Infrared Camouflage Coatings from a Cephalopod Protein

*Long Phan, Ward G. Walkup IV, David D. Ordinario, Emil Karshalev, Jonah-Micah Jocson, Anthony M. Burke, and Alon A. Gorodetsky**

Submitted to ADVANCED MATERIALS

Supporting Information

DOI: 10.1002/adma.201301472

Reconfigurable Infrared Camouflage Coatings from a Cephalopod Protein

By *Long Phan*, *Ward G. Walkup IV*, *David D. Ordinario*, *Emil Karshalev*, *Jonah-Micah Jocson*, *Anthony M. Burke*, and *Alon A. Gorodetsky**

Alon A. Gorodetsky,* Long Phan, David D. Ordinario, Emil Karshalev, Jonah-Micah Jocson, Anthony M. Burke
Department of Chemical Engineering and Materials Science
University of California, Irvine
Irvine, CA 92697 (USA)
E-mail: (alon.gorodetsky@uci.edu)

Ward G. Walkup IV
Division of Biology
California Institute of Technology
Pasadena, CA 91125 (USA)

Detailed Description of Film Thickness Calculations. The theoretical film thicknesses were calculated based on thin-film interference theory.[1] At normal incidence, the reflected light obeys the formula:

$$m * \lambda = 2 * n * d$$

where *m* is an integer that describes wavelength order, λ is the peak reflected wavelength, *n* is the refractive index of the film, and *d* is the film thickness. At non-normal incidence, a factor of $\cos(\theta_F)$ must be introduced, where θ_F is the angle in the film that can be calculated with Snell's Law. The formula above is applicable for measurements performed with an integrating sphere accessory because the angle of incidence for the integrating sphere is less than 10° from normal incidence, leading to a small deviation in the theoretical thickness (< 4 %).[1] The graphene oxide adhesion layer used in our experiments was very thin with a high refractive index, so it did not dramatically influence the calculated values. The refractive indices used for the calculations were measured at Filmetrics (San Diego, CA) using a Filmetrics F20-UV spectrometer (Figure S13).

Surface Treatment. The APTES/graphene oxide surface treatment yielded smooth RfA1 films over large areas. Alternative surface functionalization schemes, such as triethoxyoctylsilane (OTES) treatment, UV-Ozone exposure, and APTES treatment without graphene oxide, produced less reliable results.

Response of Reflectin Films to Chemical Stimuli. We screened numerous potential solvents for shifting the reflectance of RfA1 coatings. These solvents included acetic acid, ethanol, ethyl acetate, isopropanol, triethyl amine, dimethyl formamide, and dimethyl sulfoxide. Such stimuli produced small spectral shifts, with the exception of acetic acid vapor.

[1] A. Lipson, S. G. Lipson, H. Lipson, *Optical Physics*, 4th ed., Cambridge University Press, Cambridge, **2010**.

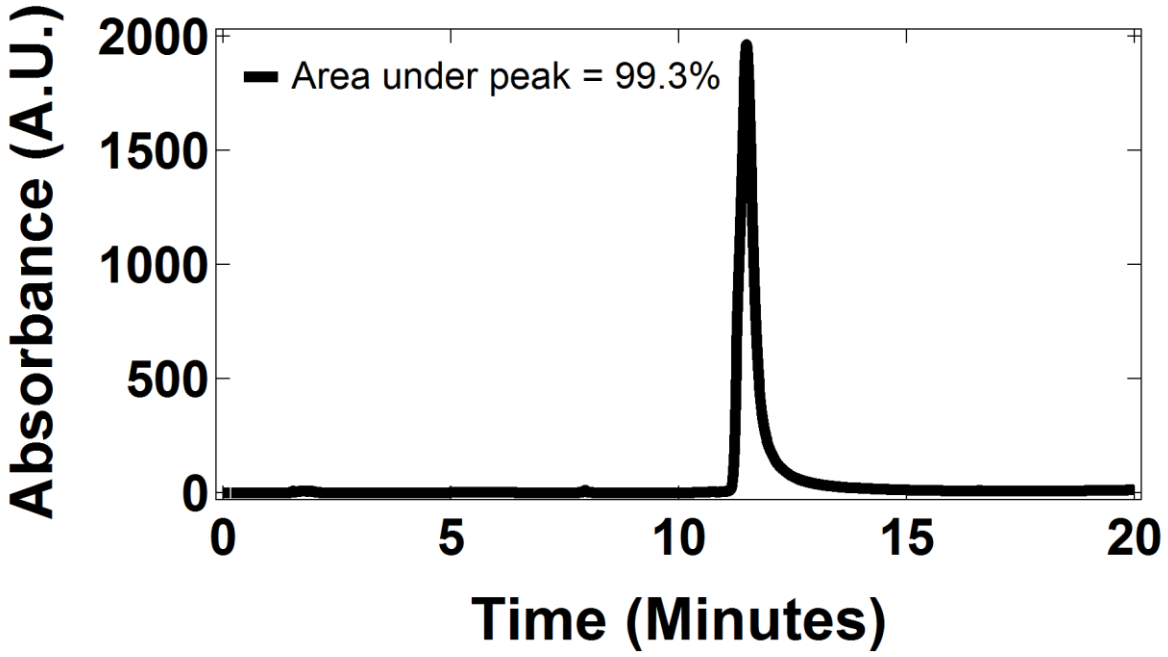


Figure S1. Typical reverse phase HPLC chromatogram of RfA1 following immobilized metal affinity chromatography. The presence of a single peak, whose area integrates to > 99 % of the total peak area represented in the chromatogram, indicates the excellent purity of our protein. The single peak also indicates that higher order aggregates are not present during chromatographic purification. RfA1 was eluted in a gradient evolved from 5 % Buffer A:95 % Buffer B to 95 % Buffer A:5 % Buffer B at a flow rate of 0.5 mL min^{-1} over 20 minutes (Buffer A: 99.9 % H_2O, 0.1% TFA; Buffer B: 95 % acetonitile, 4.9 % H_2O, 0.1% TFA).

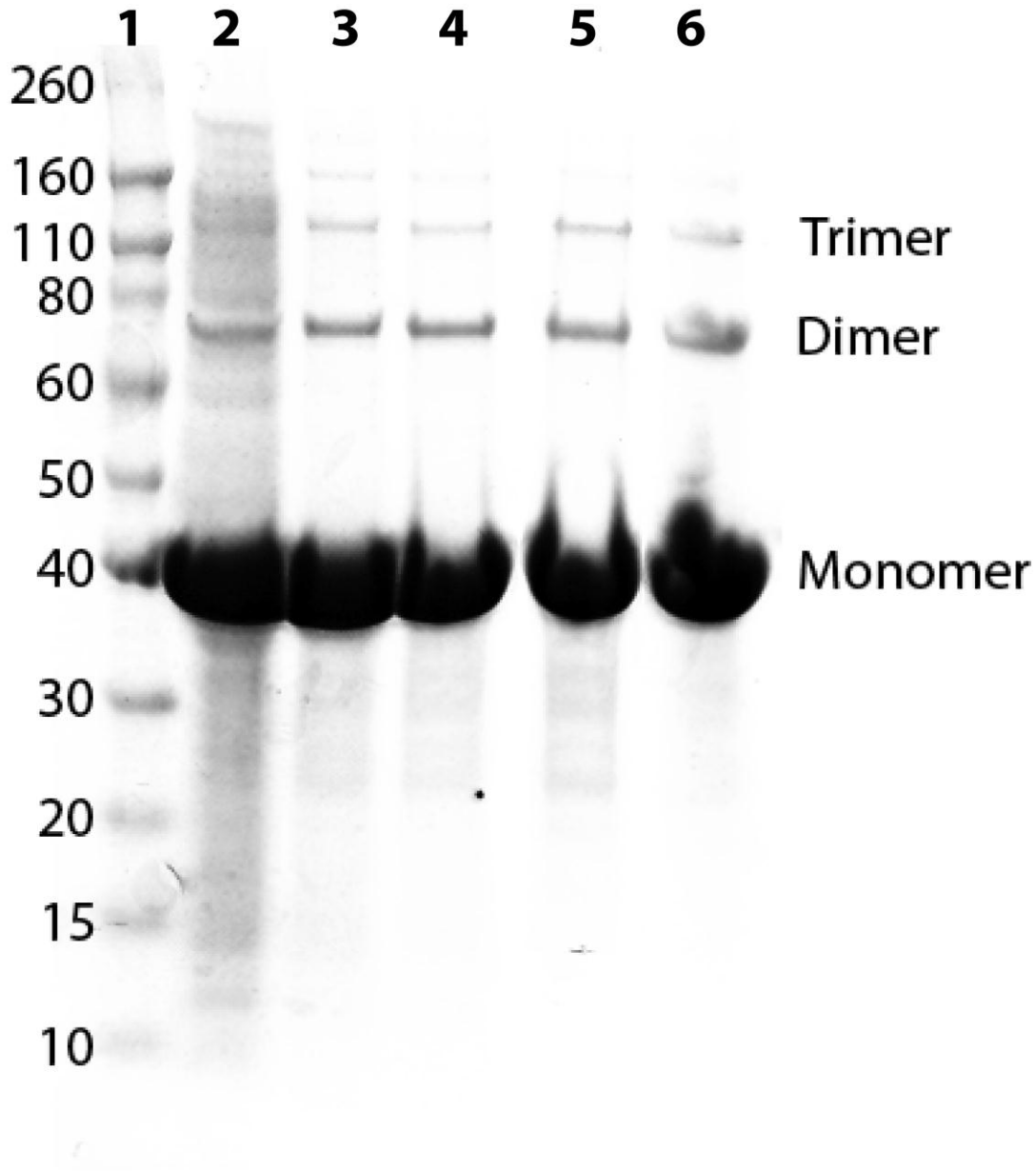


Figure S2. Analysis of the purification of RfA1 *via* denaturing gel electrophoresis. Samples of RfA1 at different stages of the purification process were analyzed by SDS-PAGE and visualized by GelCode Blue Stain Reagent (Thermo Scientific). Lane 1: Novagen Sharp protein ladder for molecular weights of 10 kDa to 260 kDa. Lane 2: Total protein fraction after cell pellet resuspension and lysis. Lane 3: Insoluble protein fraction following centrifugation. Lane 4: Inclusion body fraction after resuspension in denaturing buffer. Lane 5: Eluate from the IMAC (HisPur Cobalt Resin) gravity column. Lane 6: Eluate from C18 reverse phase HPLC column. Bands corresponding to monomeric (44,605 Da) and oligomeric RfA1 are labeled accordingly.

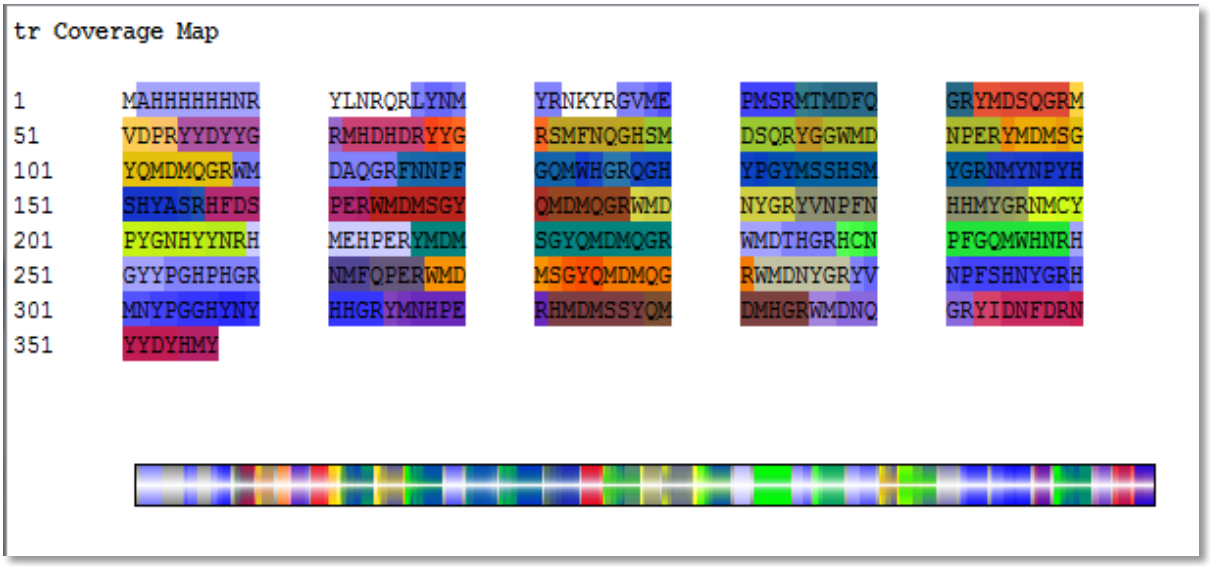


Figure S3. Peptide coverage map of RfA1 protein sequence produced by in-gel digestion of purified soluble RfA1 with porcine trypsin. The color coding corresponds to different tryptic peptide fragments. Sequence coverage exceeded 96%, definitively confirming the purified protein's identity as RfA1.

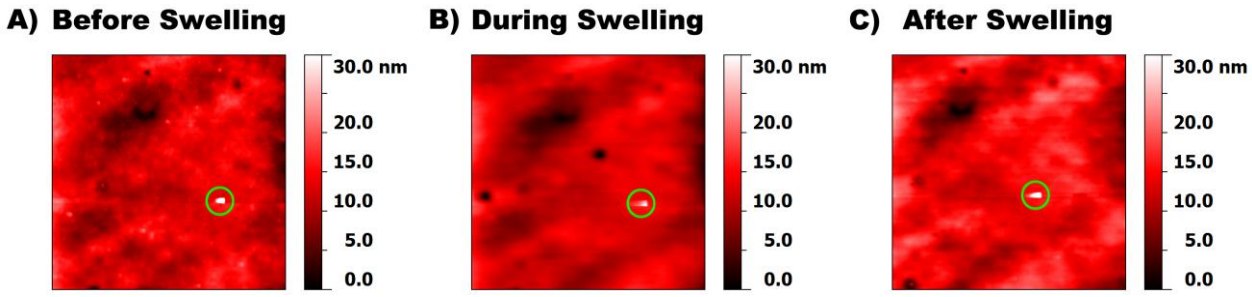


Figure S4. A) Atomic force microscopy (AFM) images of a RfA1 film A) before, B) during, and C) after exposure to acetic acid vapor. The film morphology does not exhibit any signs of degradation during and after the incubation. Note that the AFM images represent the same spot on each film as demonstrated by the green marker.

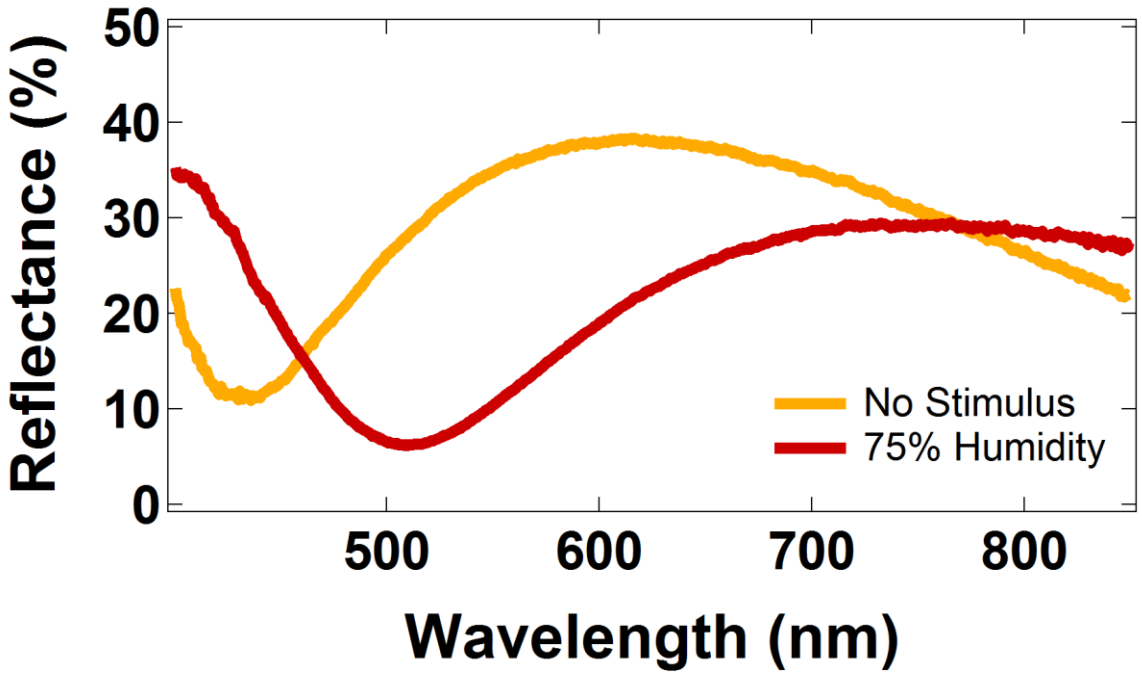


Figure S5. Reflectance spectra of an RfA1 orange film before (orange curve) and after (red curve) exposure to water vapor (relative humidity of 75%).

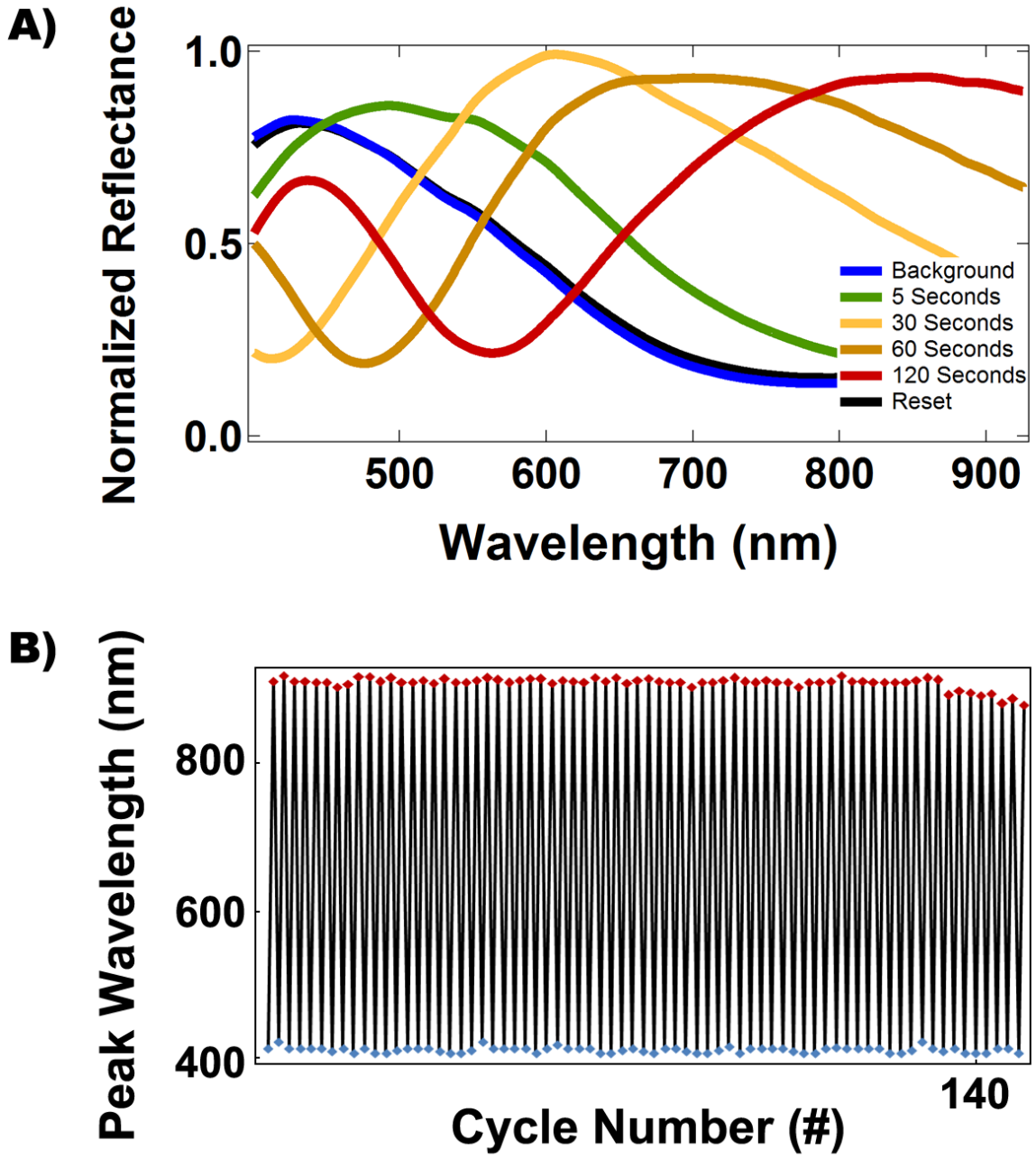


Figure S6. A) Change in the reflectance for an RfA1 film as a function of time following application of acetic acid vapor. The spectra were collected with a bare fused silica substrate as the standard and normalized with respect to the maximum reflectance value. Note that the initial reflectance spectra can be recovered within seconds by a "reset" or removal of the acetic acid vapor stimulus (the blue and black traces overlap). B) Illustration of reversible cycling of the RfA1 films between a peak reflectance of λ = 400 nm in the visible (blue markers) and a peak reflectance of λ = 900 nm in the infrared (red markers). The cycling was performed by exposure to solvent vapor from a glacial acetic acid solution. The experiment was repeated until the film demonstrated a degradation in the reflectance intensity.

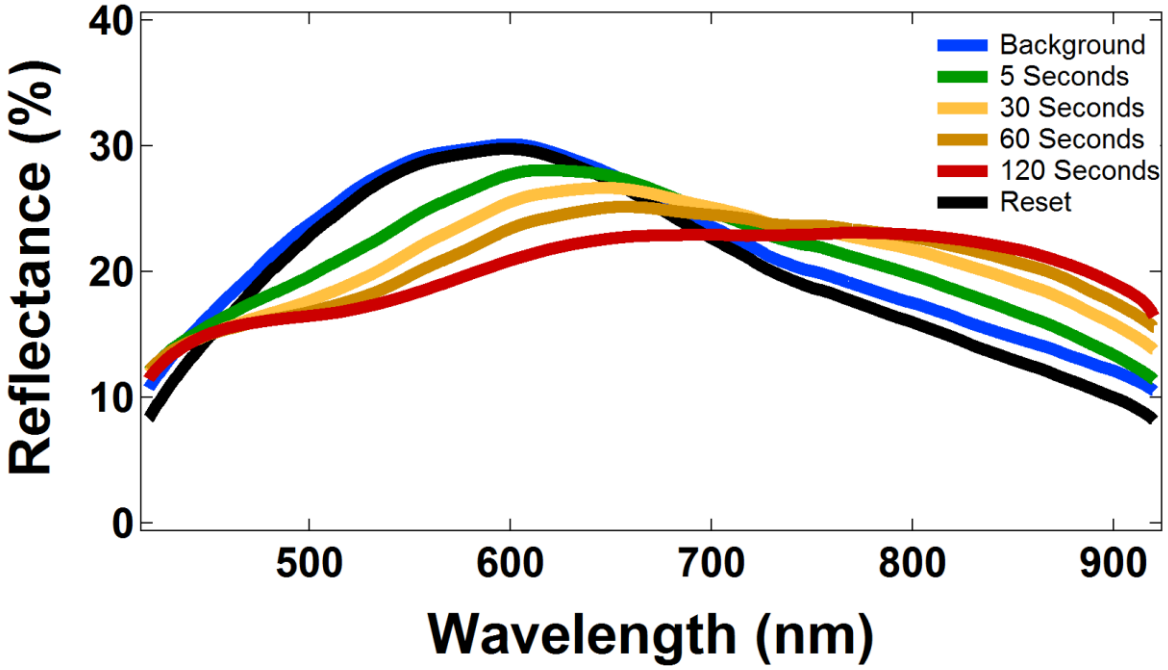


Figure S7. Reflectance spectra for a Bovine Serum Albumin (BSA) film as a function of time following application of acetic acid vapor. The film preparation and treatment were identical to those used for RfA1 films. The curves have been smoothed for clarity. Note the small peak shift.

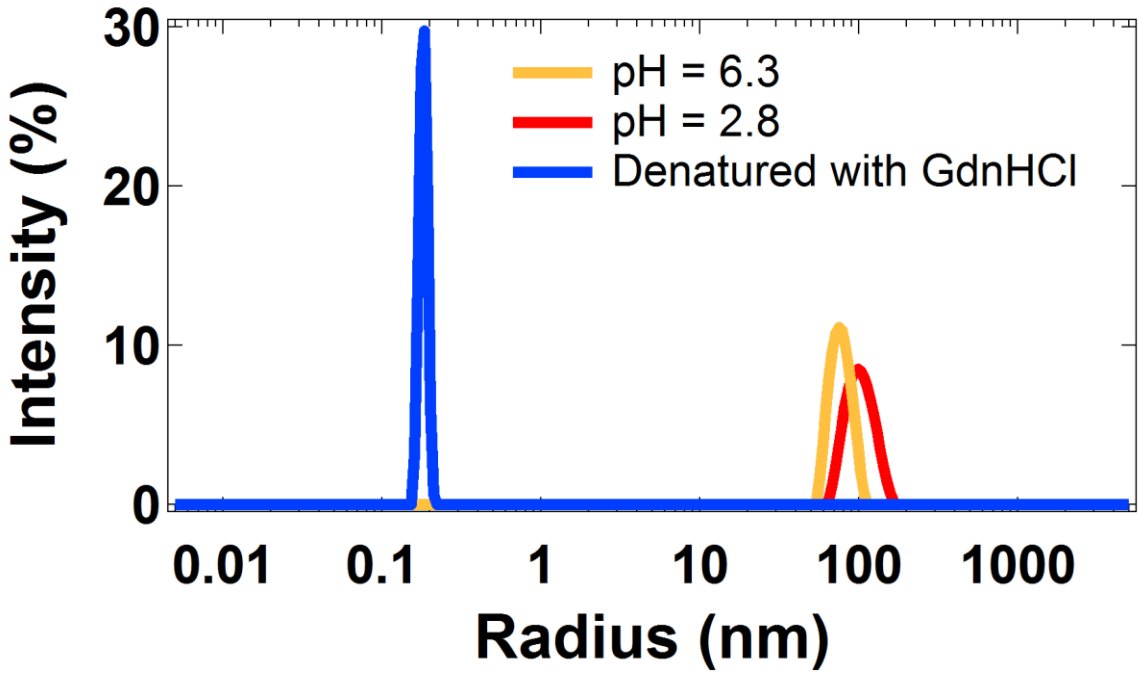


Figure S8. Dynamic light scattering spectra for RfA1 (0.5 mg/mL) at pH 6.3 (orange curve) and pH 2.8 (red curve). The measured R_H value increases from 75 (±11) nm at pH 6.3 to 99 (±14) nm at pH 2.8. The size of the aggregated RfA1 is distinct from that reported in reference 15 in the text, presumably due to differences in the protein and salt concentrations. The aggregation of RfA1 was confirmed by the introduction of guanidinium hydrochloride (GdnHCl) into the solution, which denatured RfA1 (blue curve), yielding an R_H value of < 1 nm.

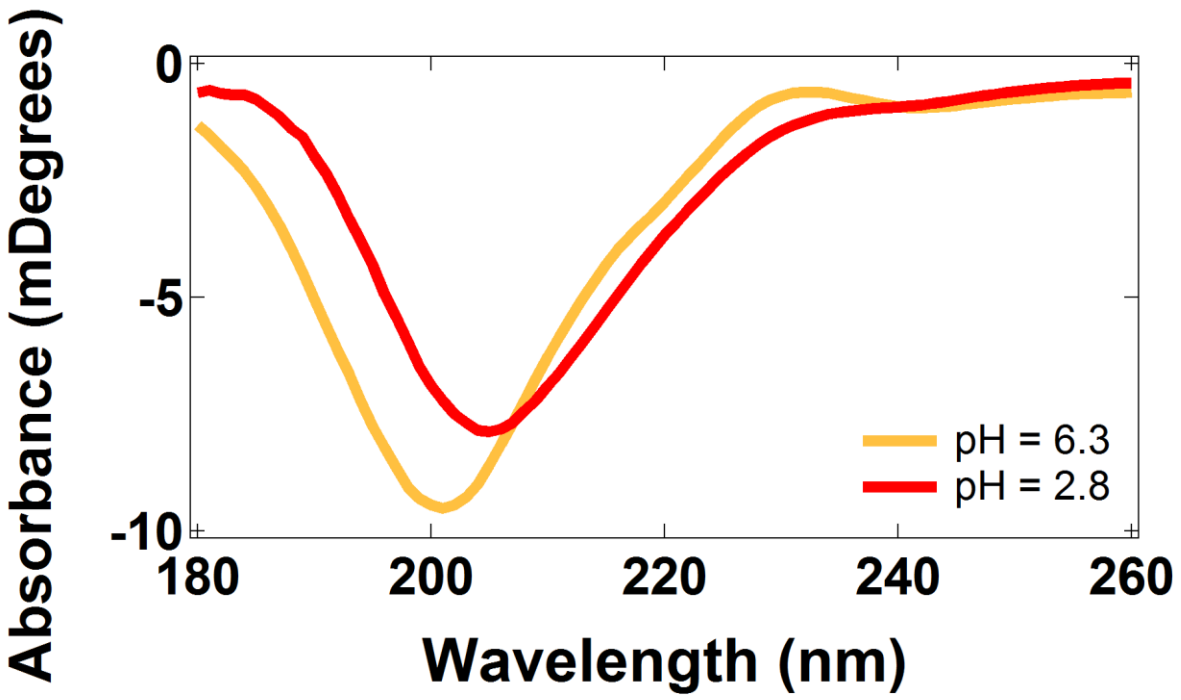


Figure S9. Circular dichroism spectra for RfA1 (~0.5 mg/mL) at pH 6.3 (orange curve) and pH 2.8 (red curve). The spectra were analyzed with DICHROWEB software (http://dichroweb.cryst.bbk.ac.uk), indicating that RfA1 possesses little to no alpha helical or beta strand secondary structure, which is in excellent agreement with previous findings (see references 13 to 17 in the text).

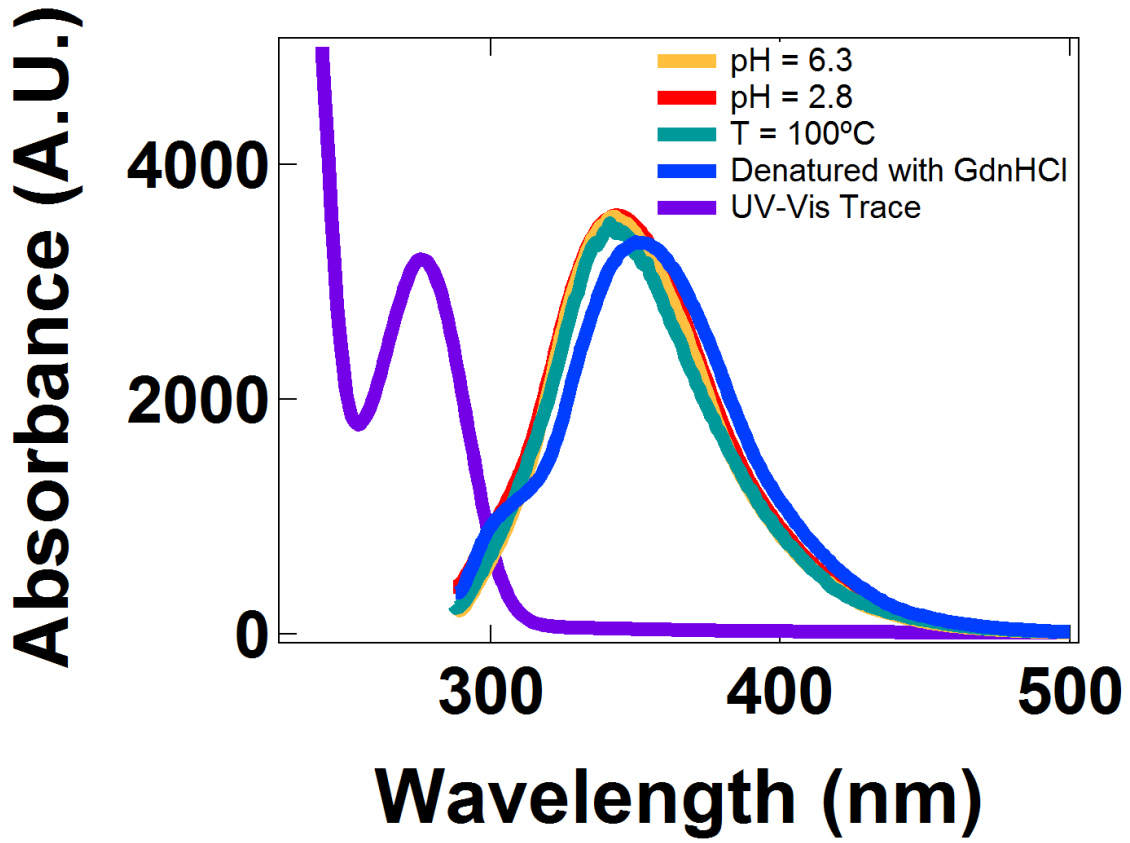


Figure S10. Fluorescence and UV-visible absorbance spectra for RfA1 (0.5 mg/mL). The absorbance spectrum (purple curve) displays a characteristic peak at λ = ~ 280 nm. The fluorescence spectra were obtained for RfA1 (0.5 mg/mL) at pH 6.3 (orange curve), pH 2.8 (red curve), and at a temperature of 100°C (green curve) (excitation wavelength of λ = 282 nm). There is virtually no change in the fluorescence spectra with a change in pH (red curve) or temperature (green curve), indicating RfA1 maintains its conformation. However, upon addition of GdnHCl (blue curve), the fluorescence spectrum maxima shifts to λ = ~ 347 nm and a new peak appears at λ = ~ 300 nm. Absorbance at λ = ~ 300 nm is characteristic of solution exposed tyrosine residues, i. e. RfA1 denaturation.

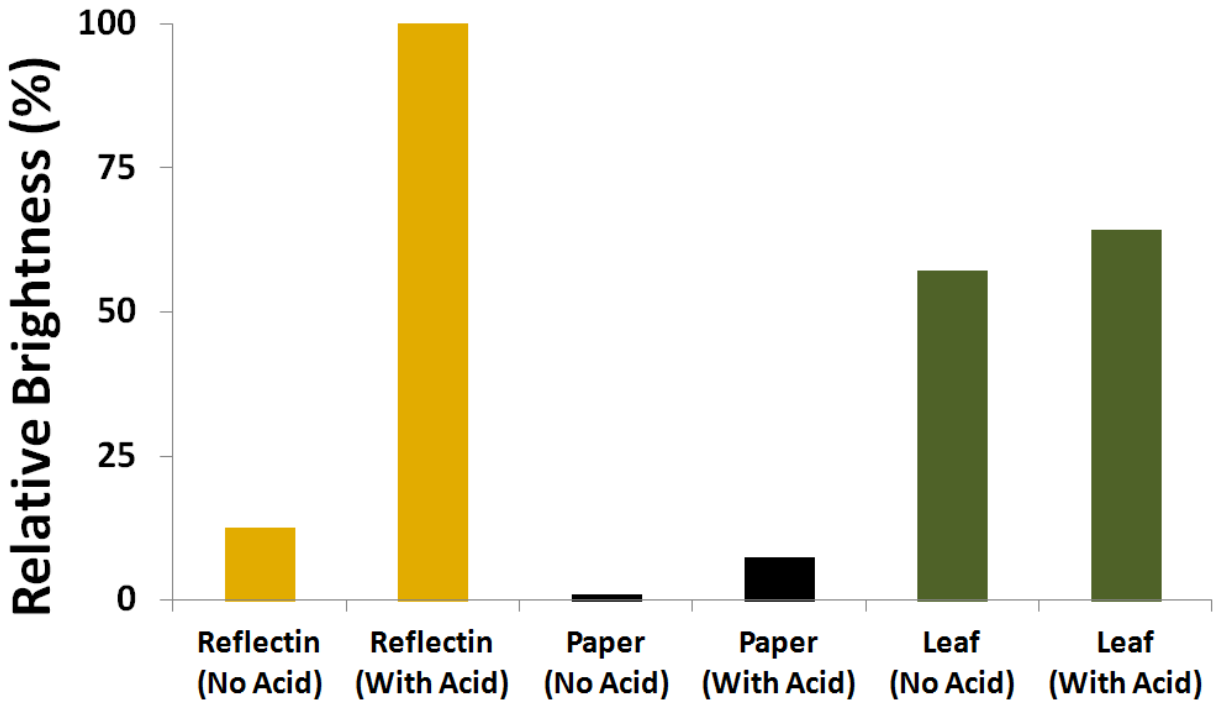


Figure S11. Integrated infrared brightness values measured for an RfA1 film, a piece of paper, and a *Buxus* plant leaf in the presence and absence of acetic acid vapor. The values were extracted by analyzing Figure 3B in the main text and normalized to the maximum brightness.

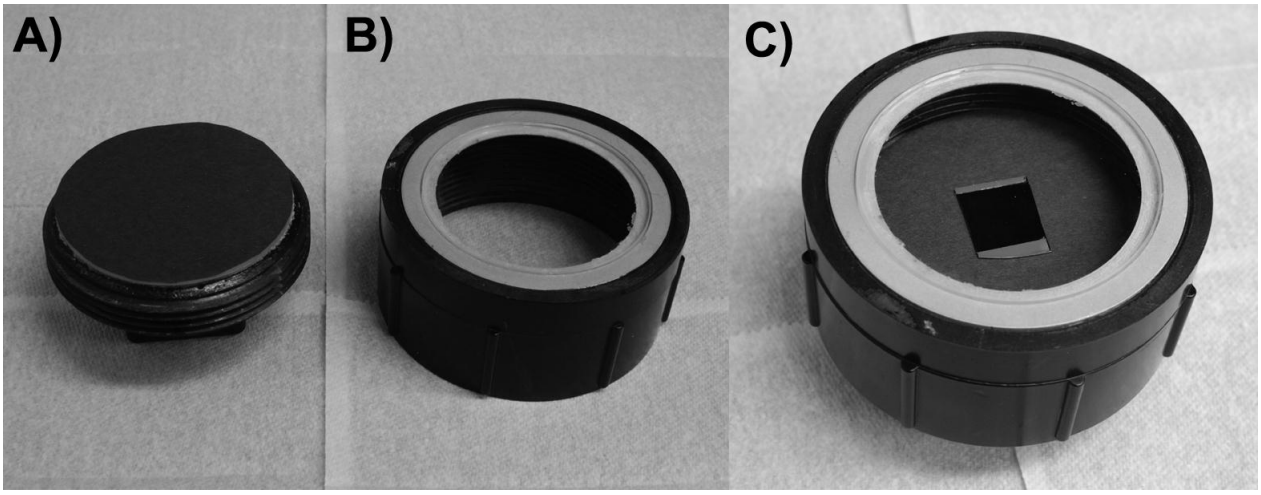


Figure S12: Pictures of the housing used for exposure of the RfA1 coated substrates to acetic acid vapor. A) The bottom part of housing. B) The top part of housing. C) The integrated housing containing an RfA1 coated substrate.

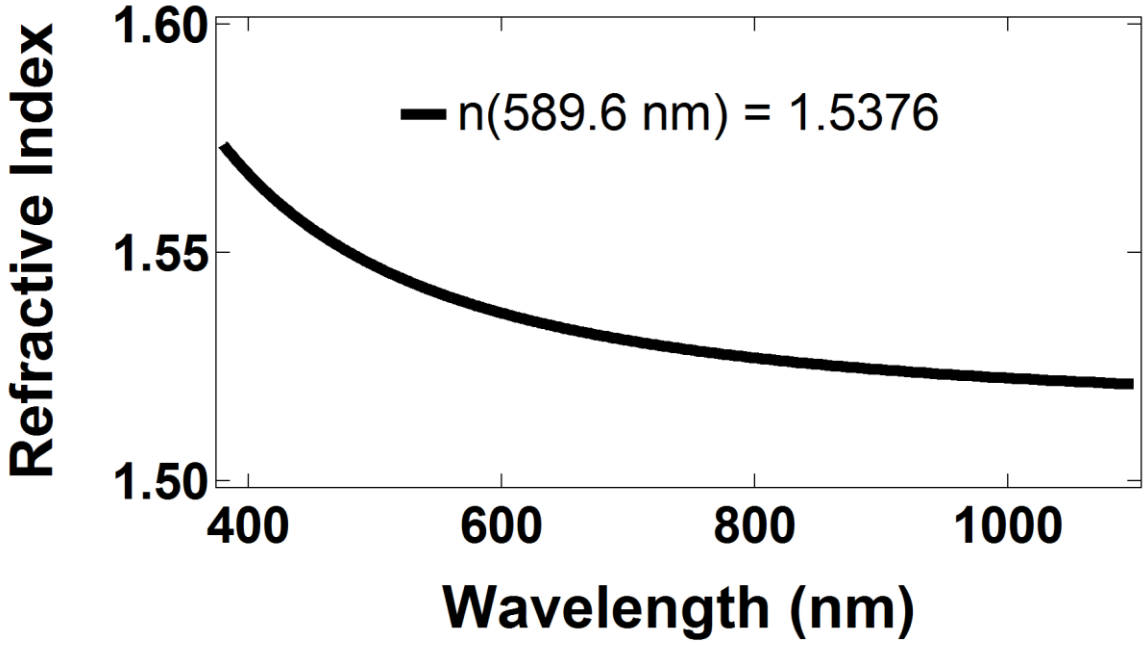


Figure S13. The refractive index of the RfA1 films measured at Filmetrics (San Diego, CA) using a Filmetrics F20-UV spectrometer with a deuterium and tungsten-halogen white light source (effective wavelength range of λ = 200 nm to λ = 1100 nm). The refractive index was measured at λ = 632.8 nm with a resolution less than 1 nm. The data was acquired normal to the substrate surface.