MOLECULAR EVOLUTION OF THE GUANYLATE KINASE DOMAIN

by

DOUGLAS PAUL ANDERSON

A DISSERTATION

Presented to the Department of Chemistry and Biochemistry
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2014

DISSERTATION APPROVAL PAGE

Student: Douglas Paul Anderson

Title: Molecular Evolution of the Guanylate Kinase Domain

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Chemistry and Biochemistry by:

Brad Nolen	Chairperson
Kenneth Prehoda	Advisor
Joseph Thornton	Core Member
Tom Stevens	Core Member
Karen Guillemin	Institutional Representative

and

J. Andrew Berglund	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2014

DISSERTATION ABSTRACT

Douglas Paul Anderson

Doctor of Philosophy

Department of Chemistry and Biochemistry

December 2014

Title: Molecular Evolution of the Guanylate Kinase Domain

The evolution of novel protein functions and protein families is a fundamental question within both evolutionary biology and biochemistry. While many gene families follow predictable patterns of molecular tinkering, many protein families exist with completely novel functions now essential. The guanylate kinase protein interaction domain (GK_{PID}) of the membrane associated guanylate kinases (MAGUK) represents a model system for the study of protein evolution in which a protein scaffolding domain has evolved from a nucleotide kinase ancestor. Here we elucidate the ancient mechanisms by which these new functions evolved by combining ancestral protein reconstruction with in vitro and cell-biological molecular experiments. We found that the GK_{PID}'s capacity to serve as a mitotic spindle-orienting scaffold evolved by duplication and divergence of an ancient guanylate kinase enzyme before the divergence of animals and choanoflagellates. Re-introducing a single historical substitution into the ancestral guanylate kinase is sufficient to abolish the ancestral enzyme activity, confer the derived scaffolding function, and establish the capacity to mediate spindle orientation in cultured cells. This substitution appears to have revealed a latent protein-binding site, rather than constructing a novel interaction interface, apparently by altering the dynamics or conformational occupancy of a hinge region that determines whether the binding site is exposed or hidden. Three further

substitutions also conveyed a measure of ligand specificity to phosphorylated Pins, which is necessary in metazoan spindle orientation pathways. These findings show how a small number of simple, ancient genetic changes caused the evolution of novel molecular functions crucial for the evolution of complex animals and laid the groundwork for an entirely new family of metazoan scaffolding proteins.

This dissertation contains previously unpublished, co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Douglas Paul Anderson

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
University of Portland, Portland, Oregon

DEGREES AWARDED:

Doctor of Philosophy, Chemistry, 2014, University of Oregon
Bachelor of Science, Chemistry, 2009, University of Portland

GRANTS, AWARDS, AND HONORS:

2014 Graduate Student Research Forum Best Poster Award
Life Sciences Category, University of Oregon, 2014

Keystone Symposia: Evolution of Phosphorylation
Travel Award, University of Oregon, 2011

Air Force ROTC Scholastic Excellence Award, University of Portland, 2006

Academic Excellence Scholarship, University of Portland, 2005

PUBLICATIONS:

Douglas P. Anderson, Dustin S. Whitney, Victor Hanson-Smith, Arielle Woznica, Nicole King, Brian F. Volkman, Joseph W. Thornton, Kenneth E. Prehoda.2014. Evolution of a new protein function required for organized multicellularity in animals. In review, Nature.

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Professors Kenneth Prehoda and Joseph Thornton for their support of this dissertation and its component projects. In addition, special thanks are due to Dr. Thornton, whose expertise and mentoring in evolutionary biology was invaluable throughout the stages of the dissertation projects. I also thank the rest of my committee members Karen Guillemin, Tom Stevens and Brad Nolen as well as Will Campodonico.

I wish to dedicate this dissertation to my loving wife Jeneva whose enduring support has made this all possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Protein-protein interactions are an essential element of living systems and are essential for proper cellular functions. While some protein-protein interactions are non-specific, the many protein-protein interactions are highly specific and tightly regulated (1). Taken together, the sum of protein-protein interactions creates a dense web of interactions within the cell responsible for a wide array of cellular functions and behavior. Although many protein interactions occur in addition to a proteins primary function, many families of proteins are present within the cell whose sole function is to mediate protein-protein interactions, called scaffolding proteins. Scaffolding proteins allow the interaction between many protein partners and the potential for organized supra-molecular complexes such as seen in the metazoan synapse (2). Research in the past ten years has begun to highlight the importance of such scaffolding molecules even beyond cellular phenotypes where alterations to a scaffolding protein can have organismal effects in behavior due to the disruption of complex protein complexes involved in development and neurophysiology (3).

Membrane-Associated Guanylate Kinases (MAGUKS)

One family dedicated protein-protein interaction scaffolds is the Membrane Associated Guanylate Kinases (MAGUKs). The MAGUKs are a functionally diverse, family, involved in scaffolding for many complex cellular phenotypes such as apoptosis, adhesion junctions, pre and post-synaptic density organization, and mitotic spindle orientation (4). MAGUK family proteins are multi-domain proteins, generally consisting of at least three different classes of protein-protein interaction domains (5). A typical

MAGUK protein will consist of 1-3 PDZ domains, a small peptide-ligand recognizing domain, a single Src Homology 3 (SH3) domain which also recognizes small peptide sequences, and an C-terminal guanylate kinase-like (GK) domain for which the MAGUKs are named (Figure 1). While various MAGUK family members may also possess additional domains, all MAGUKs possess at the minimum a PDZ-SH3-GK domain organization and are defined in part by this domain architecture (6).

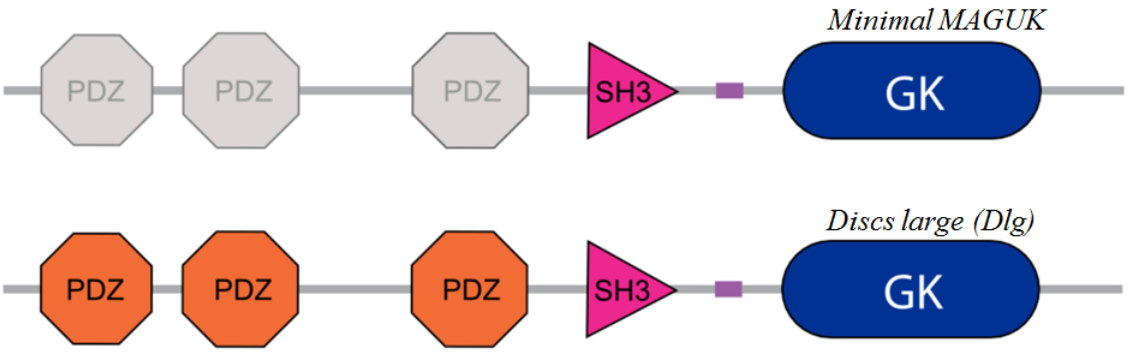


Figure 1. Domain architecture of the Membrane Associated Guanylate Kinases (MAGUKs). A minimal MAGUK (top) always consists of a C-terminal GK domain (blue) with a nearby SH3 domain (pink) and is usually accompanied by 1-3 PDZ domains (orange, gray). Additional domains specific to various MAGUK families are usually located in the N-terminus. Discs large (Dlg,bottom) is shown as an example.

While the majority of the MAGUK protein interactions are mediated by the well-studied PDZ and SH3 domains, the C-terminal GK domain is less understood. Unlike PDZ or SH3 domains which are ubiquitous and found in many other protein families, the GK domain is unique to the MAGUK family. The GK domain itself was first discovered in the early 90's (7) and was identified and named by sequence homology to the Guanylate Kinases, a nucleotide kinase responsible for phosphorylating GMP with ATP (8). In these early days, it was believed that like the guanylate kinases, the MAGUK GK domain was also capable of catalysis and that the protein served to localize enzymatic activity to precise cellular locations (9). A few well known MAGUKs known to localize to the post-synaptic

density helped give rise to this notion and are the source of the MAGUK family name due to the membrane localization patterns and the eponymous GK domain (10). Functional studies in the late 90's soon dispelled the notion of a catalytically active GK domain (8), finding that not only was the GK domain inactive but could not be reactivated through mutation despite possessing a seemingly intact catalytic core (11).

The function of the GK domain was later identified in the mid 2000's with a combination of approaches. Multiple genetic, crystallography and biochemical experiments showed that the GK domains of the Dlg were in fact binding protein ligands and not nucleotides as previously identified . Crystallographic studies of the GK domain from Discs large (Dlg) have identified multiple protein ligands bound involved in spindle orientation pathways as well as neuronal synaptic ligands (12,13). Studies of non Dlg-family GK domains have also shown protein ligand binding, suggesting protein binding may be a general feature of GK domain function (4) . Of the many GK domains, the Dlg family GK domain (Dlg-GK) remains the best studied, and will be the focus of this dissertation.

Guanylate Kinase Domains of Dlg Proteins

Although the Dlg family GK domain is involved in many cellular pathways and protein interactions (14,15), its most well characterized functional role lies in its participation in orienting the mitotic spindle (16). In this role, the GK domain serves a crucial role as a scaffolding protein linking cortically associated proteins to the mitotic spindle through microtubule-associated protein interactions. In *Drosophila melanogaster* models of spindle orientation, this is accomplished by binding cortically localized Partner of Inscuteable (Pins), an essential spindle orientation and cell polarity protein, and the

microtubule-associated motor protein KHC-73 (17). Forming this Pins/GK/KHC-73 complex serves to connect the mitotic spindle to a specific axis of the cell, biasing the direction of the mitotic spindle towards the localization of Pins during mitosis (Figure 2, ref. 18). Although additional pathways mediated by Pins can build upon this initial spindle orientation capability, the Pins/GK/KHC-73 complex as a necessary first pathway which additional pathways require to function (17). Compromising the scaffolding activity of the GK domain in this pathway or disrupting the other complex members also leads to impaired spindle orientation, defects in cell polarity and differentiation among other phenotypes (19,20,21).

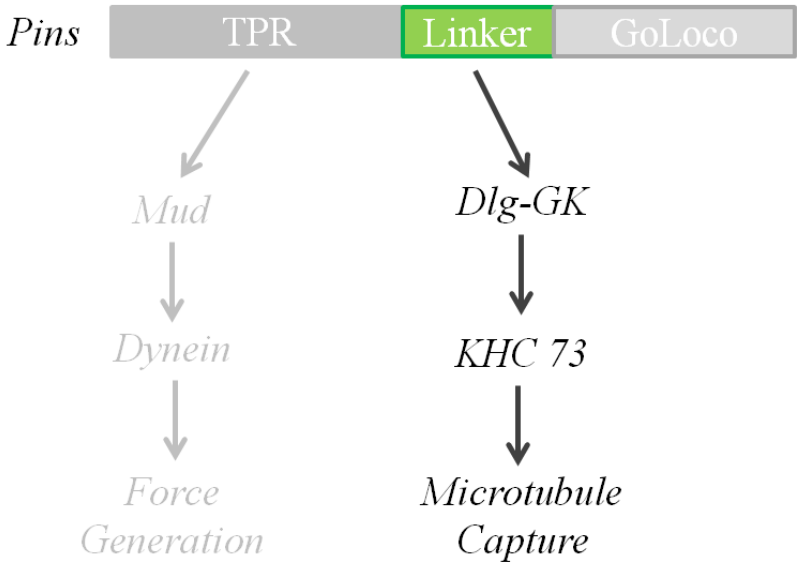


Figure 2. Genetic pathway of spindle orientation in the *Drosophila* neuroblast. The GK domain of Dlg binds to the linker region of the protein Partner of Inscuteable (Pins, green). Further protein-protein interactions between the GK domain and the astral microtubular associated motor protein KHC-73 are believed to capture the mitotic spindle, biasing the spindle towards the localization of Pins. Additional Pins pathways (grey) further refine spindle orientation but require the Pins / GK pathway to function.

More recent work have further elucidated the functional mechanisms behind the Dlg-GK domains scaffolding function through crystallographic means, uncovering the

biophysical mechanism behind the GK domains protein-binding function (12,13). In one study, the Dlg-GK domain was co-crystallized in complex with a Pins-peptide ligand identified from previous genetic studies. Surprisingly, this crystal showed the Pins ligand interacting with a remarkably intact guanosine-binding subdomain, once believed to bind GMP (Figure 3, 12). More surprisingly, it was found that the Pins ligand inserts a phosphorylated serine into a negatively charged knob very similar to that of the guanylate kinase enzyme. Comparisons of the two structures show nearly identical secondary and tertiary structure, where the guanylate kinase guanosine binding domains are almost completely conserved in the protein-binding GK domain. These studies gave further evidence to the growing hypothesis that GK domains are likely derived from an ancient guanylate kinase ancestor which has been co-opted to bind protein ligands with little structural modification.

Further biochemical work among extant GK enzymes and domains has even suggested potential mutations which may have been involved in the neofunctionalization between the hypothesized guanylate kinase ancestors and the GK domains. In a recent study, a conserved proline residue found in most GK domains including Dlg family GK domains substituted was inserted into the same location in native guanylate kinases and assayed for potential effects (22). Surprisingly, the authors found that when the guanylate kinase serine identity was mutated to the GK domain proline identity most enzymatic activity was lost while gaining the ability to act as GK domains in spindle orientation assays and binding Pins with low affinity. This mutation appeared to slow the guanylate kinases' enzymatic motions which were correlated with the gain of protein binding and spindle orientation capability. From this study, the authors suggest that this mutation (ser to

pro) may have been an evolutionarily relevant neofunctionalizing substitution that likely contributed to GK domain evolution.

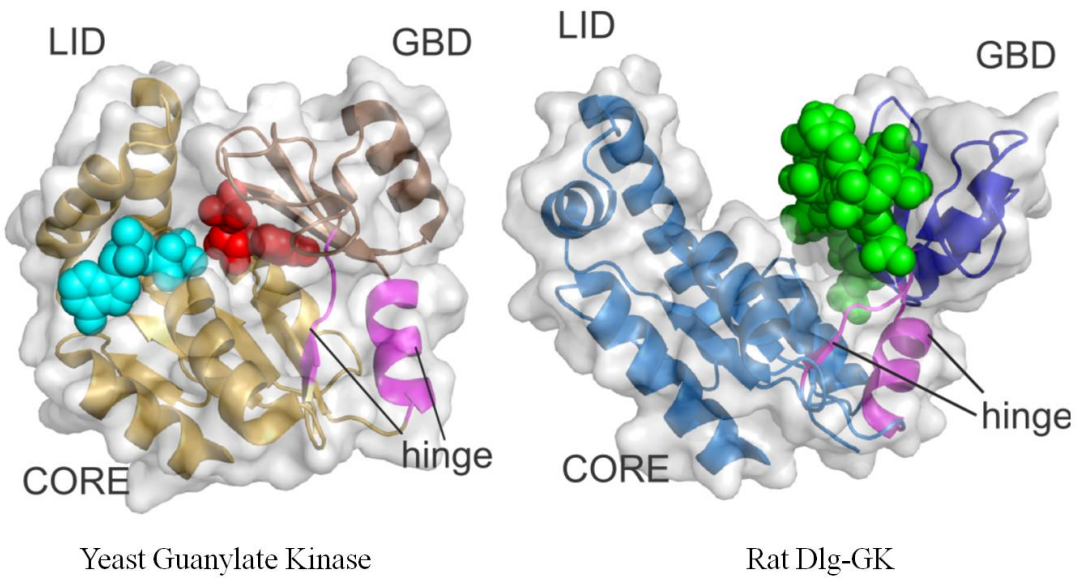


Figure 3. Comparison of guanylate kinases and GK domains. Extant yeast guanylate kinase is shown (left, brown, PDB:1GKY) with bound ATP (cyan spheres) in the LID subdomain, GMP (red spheres) in the GBD subdomain. GBD and LID subdomains are connected to the central CORE subdomain by flexible hinge regions (pink). GK domains (right, blue, PDB:3UAT) possess the same subdomain structure but bind the much large protein ligand Pins (green spheres) where GMP was previously bound.

Phylogenetics of the GK Domain

Alongside these structural lines of evidence, phylogenetic studies of the GK domain from all MAGUK family members have also begun postulating a direct evolutionary origin from the guanylate kinases (Figure 4, 23). Evidence for this hypothesis lies in the species distribution of the guanylate kinases themselves compared to that of MAGUK GK domains. While all life forms on Earth possess at least one guanylate kinase, GK domains are limited to metazoa and choanoflagellates (24). Such a distribution suggests that the MAGUK GK domain is the result of a gene duplication and divergence of an ancestral

guanylate kinase somewhere before metazoa. Further gene duplication and differentiation would then have given rise to the various MAGUK family. Two phylogenetic studies (23, 24) have provided support for this hypothesis by constructing a well-supported gene phylogeny using the guanylate kinases as an outgroup for the MAGUK GKs. Though the phylogenetic support for the duplication and divergence of the GK domains from an ancient guanylate kinase duplication event is evidence, many questions remain. Beyond the duplication and divergence itself, how did an ancient guanylate kinase evolve protein binding ability? More specifically, how did the Dlg family functionality of spindle orientation evolve? When was guanylate kinase activity lost? What are the genetic changes that are necessary for such a transition? What are the underlying biophysical mechanisms behind these genetic changes? Further questions remain specific to the Dlg-GK family, such as when did the Dlg-GK spindle orientation function evolve in MAGUK GK evolution? Is GK spindle orientation function a basal function or a more recently evolved phenotype?

Prior to this dissertation, only suggestive conclusions could be made based on comparisons of extant guanylate kinases and domains. As discussed earlier,biochemical experiments with *S. cerevisiae* guanylate kinase identified that when a strictly conserved serine residue in the guanylate kinase hinge region, a region necessary for catalytic movements, was mutated to the GK domain proline identity was sufficient to cause a loss of enzyme activity and a gain of protein binding and spindle orienting ability (22).

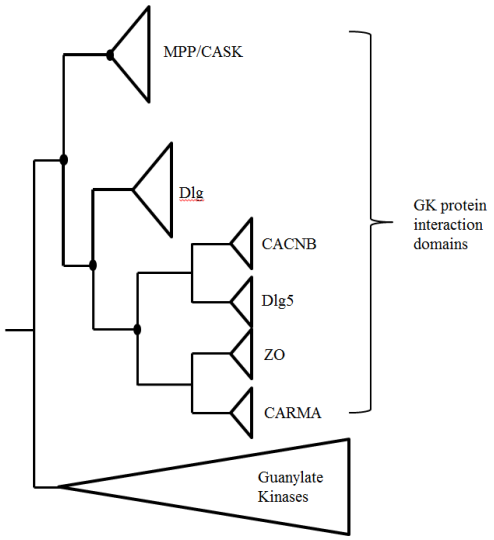


Figure 4. Speculated evolutionary origin of the GK protein interaction domains. All GK domains (bracketed) are speculated to have evolved from a common ancestor, which itself is the result of an ancient duplication and divergence event from the guanylate kinase enzyme (adapted from Mendoza et al, see reference 23).

Although suggestive of a potentially evolutionary relevant neofunctionalizing historical substitution, such extant comparisons suffer from the limitations of horizontal evolutionary analysis. Horizontal comparisons between extant taxa are at best limited, as one cannot differentiate between historically relevant identities versus neutral mutations or mutations involved in different functions. In order to truly investigate past molecular evolution, vertical evolutionary analysis studying the ancient sequences themselves is necessary to identify historically relevant substitutions. Such vertical analysis, while impossible in the recent past, can be done using the technique of ancestral protein reconstruction (APR) where ancient sequence identities are reconstructed using statistical methods applied with a phylogenetic tree (25). Reconstructed protein sequences can then be synthesized and expressed and functionally characterized to determine both the historical changes necessary for functional evolution as well as the investigation of the biophysical mechanisms behind

them. Because direct ancestral protein states can be assayed and historical substitutions directly analyzed for functional effects, APR represents a true vertical analysis allowing investigation of the causative substitutions in protein evolution and has been used successfully in many other studies to investigate protein evolution (25, 26)

In this dissertation, I will discuss my efforts to elucidate the molecular evolution of the Dlg family GK domain from a hypothesized guanylate kinase ancestor using vertical evolutionary analysis by way of ancestral protein reconstruction. The following chapters will overview both the establishment of GK domain ancestral protein reconstruction, initial functional characterization and the identification of key historical residues partially responsible for the evolution from an ancient guanylate kinase ancestor to that of a protein-binding GK domain like protein.

The following chapter contains co-authored material authored by myself and collaborator Arielle Woznica (King lab, UC Berkeley), who performed spindle orientation cell culture experiments in the choanoflagellate *Salpingoeca rosetta.*

CHAPTER II

EVOLUTION OF A PROTEIN FUNCTION REQUIRED FOR ORGANIZED MULTICELLULARITY IN ANIMALS

This chapter contains co-authored material. Arielle Woznica (King lab, UC Berkeley) performed choanoflagellate *Salpingoeca rosetta* spindle orientation experiments (Figure 9D) only.

Introduction

The evolution of organized multicellular animals from single-celled ancestors was one of the most important transitions in the history of life (27,28), but the genetic and molecular mechanisms for this transition are poorly understood. To form tissues and differentiate cell types, the mitotic spindle must be spatially oriented during cell division (29), a process mediated by the GK protein interaction domain (GK_{PID}), a scaffolding domain that links microtubular motor proteins to the cell cortex during mitosis (17). Here we elucidate the ancient mechanisms by which these functions evolved by combining ancestral protein reconstruction with in vitro and cell-biological molecular experiments. We found that the GK_{PID}'s capacity to serve as a spindle-orienting scaffold evolved by duplication and divergence of an ancient guanylate kinase enzyme, before the divergence of animals and choanoflagellates and therefore before the advent of true multicellularity. Re-introducing a single historical substitution into the ancestral guanylate kinase is sufficient to abolish the ancestral enzyme activity, confer the derived scaffolding function, and establish the capacity to mediate spindle orientation in cultured cells. This substitution

appears to have revealed a latent protein-binding site, rather than constructing a novel interaction interface, apparently by altering the dynamics or conformational occupancy of a hinge region that determines whether the binding site is exposed or hidden. These findings show how a simple, ancient genetic change caused the evolution of a novel molecular function crucial for the evolution of complex animals.

The advent of organized multicellular animals – organisms composed of differentiated cell types spatially organized into tissues – is one of the most important but least understood transition in the history of life (30,31). Phylogenetic analyses have established that many protein families now involved in cell adhesion, signal transduction, and differentiation first appeared in genomes before the evolution of true multicellularity (32-35). Virtually nothing is known, however, concerning the molecular mechanisms by which these proteins' functions evolved. These events happened in the very deep past, so horizontal comparisons between extant species cannot reveal the historical changes in protein sequence, function, or biophysical properties that caused them. Vertical evolutionary analysis using ancestral protein reconstruction (APR) – phylogenetic inference of ancestral sequences followed by gene synthesis, genetic manipulation, and experimental characterization – has proven to be an effective strategy for answering these questions (25,26). Here we apply APR to dissect the mechanisms for the evolution of a new protein function crucial to the emergence of tissue differentiation and organization in animals.

Spindle orientation is essential for organized multicellularity. To generate and maintain tissues, the plane in which cells divide must be coordinated by orienting the mitotic spindle relative to surrounding cells (36,18) Cells that orient the mitotic spindle

parallel to the epithelial plane, for example, expand the tissue; those that orient it orthogonally to the plane escape the epithelium, as in epithelial to mesenchymal transitions during development (17,18,36) Spindle orientation is also necessary for differentiation of many cell types, because it allows the dividing cell to segregate fate determinants asymmetrically among its daughters (18,37,38) Robust positioning of the mitotic spindle in animals is mediated by the scaffolding activity of the protein Discs large (Dlg), a multidomain protein that uses its GK_{PID} to bind to the Partner of Inscuteable (Pins) protein -- which is specifically localized in the cell cortex -- and to microtubule-associated motor proteins, such as KHC-73 (Fig. 6A, refs. 16,17,39). Compromising Dlg's GK_{PID} or other components of the Dlg-Pins-KHC73 complex leads to impaired spindle orientation, defects in cell polarity and differentiation, tumorigenesis, and developmental abnormalities, including epithelial and neuronal neoplasia and other failures of tissue organization (17,36,39-42)

Little is known concerning the evolution of spindle orientation or of Dlg's spindle-orienting functions. Dlg is a member of a larger family of membrane-associated multidomain proteins, all of which contain a GK_{PID} and which mediate protein-protein interactions important to cell adhesion, neural synapse organization, spindle orientation, and other functions (4). The GK_{PID} is found only in metazoa and choanoflagellates (23), but it shares strong sequence and structural homology with the guanylate kinase (gk) enzymes, which are common to all life on earth and which regulate nucleotide homeostasis by catalyzing the transfer of phosphate groups from ATP to GMP (8). These observations suggest that the GK_{PID} of Dlg and related proteins may have evolved from an ancient gk enzyme, at some time near the base of metazoans. Understanding the evolution of the

GK_{PID}'s scaffolding/spindle orientation functions is therefore important for understanding the evolution of animal complexity. It also serves as a model for the evolution of entirely new molecular functions; most studies of protein evolution to date have focused on more subtle shifts in function, such as changes in relative ligand preference, in allosteric regulation, or in quantitative measures of activity (26). To reveal how the GK_{PID}'s novel scaffolding/spindle-orienting functions evolved, we used APR to reconstruct the sequences of ancestral members of the protein family that contains the gk enzymes and GK_{PID}s, trace their functional evolution through time, and dissect the genetic and biophysical mechanisms that mediated the evolution of the GK_{PID}'s new functions.

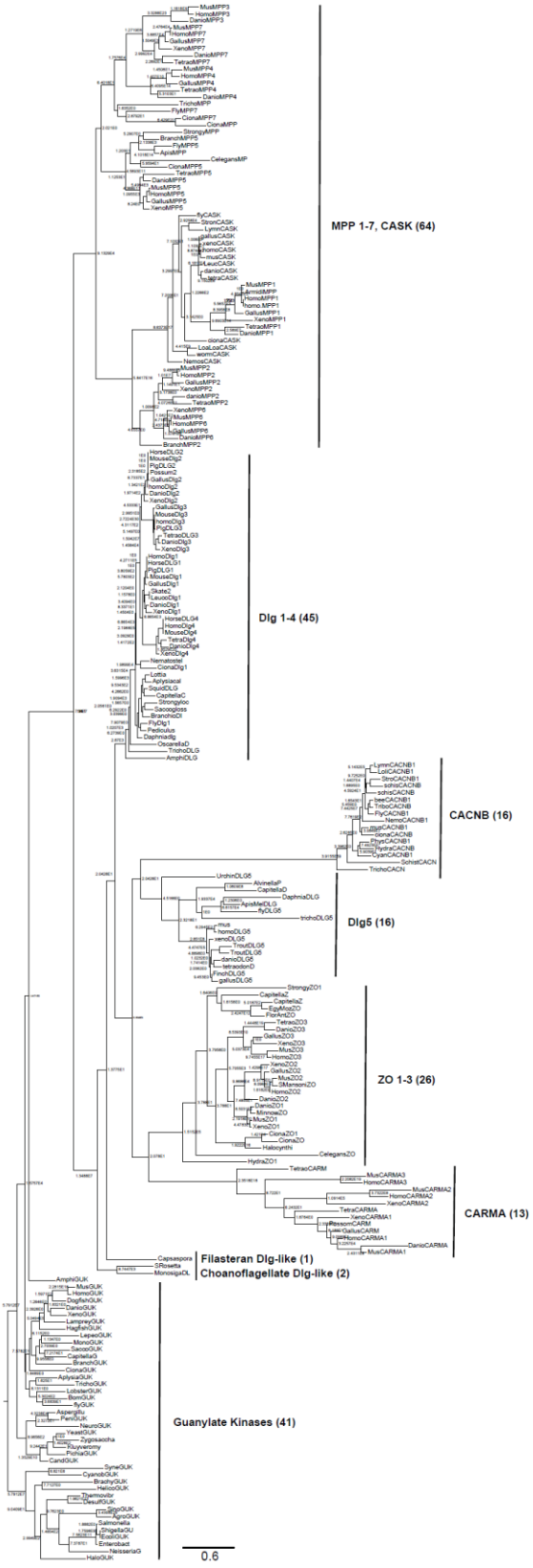
Results

We first inferred the phylogeny of the gk enzyme/GK_{PID} protein family using maximum likelihood phylogenetics and the amino acid sequences of 224 family members (Fig. 5 for full phylogeny; Fig6B). The tree was rooted using bacterial gk enzymes – the only family members present in prokaryotes – as an outgroup. All the GK_{PID}s of animals cluster together as a monophyletic group, with the gk enzymes forming a paraphyletic set of basal lineages. Dlg and related proteins comprise one of two major groups within the clade of GK_{PID}s. Like animals, choanoflagellate and filasterean species – the two unicellular taxa most closely related to Metazoa – contain a gk enzyme and a GK_{PID}, and the latter proteins occupy a well-supported basal position sister to the metazoan Dlg-containing clade. This topology indicates that the GK_{PID} was generated by $_{dup}$lication of an ancient gk enzyme before the last common ancestor of Filozoa (animals+choanoflagellates+filasterea) but after the split of Filozoa from the lineage leading to fungi and related protists, which contain no GK_{PID}s. Further gene duplications

within the animals produced the numerous proteins that now contain GK_{PID}s (Fig. 6B, ref. 23).

To understand how and when the scaffolding and spindle-orienting functions of the GK_{PID} evolved, we used maximum likelihood phylogenetics to reconstruct ancestral sequences at critical nodes on the tree. We focused on two key ancestral nodes: Anc-$GK1_{PID}$, which represents the last common ancestral protein from which all GK_{PID}s in Filozoa descend, and Anc-gk_{dup}, which existed just before the gene duplication that split the gk enzymes from the GK_{PID}s. Anc-gk_{dup} was reconstructed with strong confidence (mean posterior probability per site 0.94, with only 20 ambiguously reconstructed sites, defined as having a second plausible reconstruction with PP > 0.20); Anc-$GK1_{PID}$ was reconstructed with lower confidence (mean PP = 0.77, and 51 ambiguous sites, see Fig. 7A-B and Appendix Figures A1 and A2). Anc-gk_{dup} was most similar in sequence to extant gk enzymes, and Anc-$GK1_{PID}$ and was more similar to the GK_{PID}s than to any gk enzyme (see Figure 7C). This pattern, together with the nested phylogenetic position of the GK_{PID}s relative to gk enzymes, led us to hypothesize that that the functional transition from enzyme activity to protein-binding/spindle orienting functions occurred during the ancient interval between Anc-gk_{dup} and Anc-$GK1_{PID}$.

Figure 5 (next page). Complete phylogeny of 224 guanylate kinase enzyme and GK_{PID}s. Nodes are labeled with approximate likelihood ratio supports, and branch lengths are in substitutions per site (see scale bar). The tree is rooted on the bacterial GK enzymes. See Appendix Figure A3 for species and accessions used. Major paralogs in the GK_{PID} family, with the number of sequences included in each, are labeled.


MPP 1-7, CASK (64)
Dlg 1-4 (45)
CACNB (16)
Dlg5 (16)
ZO 1-3 (26)
CARMA (13)
Filasteran Dlg-like (1)
Choanoflagellate Dlg-like (2)
Guanylate Kinases (41)
0.6

To test this hypothesis experimentally, we synthesized DNA sequences coding for the inferred sequences of these two proteins, expressed and purified them, and characterized their functions by 1) measuring guanylate kinase activity using an in vitro coupled-enzyme assay to track production of ADP from ATP, 2) assessing affinity for a labeled Pins peptide using fluorescence anisotropy assays, and 3) characterizing spindle-orienting function using an assay of mitotic spindle geometry in cultured cells transfected with a GK domain of interest. In the latter assay, Drosophila S2 cells are transfected with Pins fused to the cell-adhesion protein Echinoid, which causes a crescent of Pins to localize at the area of contact between adjacent cells; if the GK_{PID} from a functional Dlg-like protein is cotransfected, GK's interaction with Pins causes the spindle to align during mitosis at a right angle to the crescent, along the axis between the two cells.4

As predicted, we found that the progenitor Anc-gk_{dup} protein is an active guanylate kinase enzyme, with a Michaelis constant (K_m) comparable to that of the human enzyme, albeit with a slower k_{cat} (Fig. 5C). Anc-gk_{dup} displayed no Pins binding and did not orient the mitotic spindle in living cells (Fig. 5D,G). In contrast, its descendant Anc-$GK1_{PID}$ had no detectable guanylate kinase activity, but it bound Pins with moderate affinity and was highly effective in orienting the mitotic spindle in cell culture (Figs. 5C, 1D, 1H). Corroborating this finding, we also reconstructed Anc-$GK2_{PID}$ – the more recent progenitor of all Dlg proteins in metazoans – and found that it too orients the mitotic spindle and binds Pins with high affinity (Figure 8).

To determine whether these results are robust to uncertainty about the inferred ancestral sequence, we synthesized alternate versions of both Anc-gk_{dup} and Anc-$GK1_{PID}$ that contained all plausible alternative amino acid states (defined as those with posterior

probability > 0.20). These sequences contain more differences from the ML reconstruction than the expected number of errors in the ML sequence and represent the far edge of the cloud of plausible ancestral reconstructions (see Appendix Figures A1 and A2). When tested experimentally, the alternative version of Anc-gk_{dup}, like the ML reconstruction, was an active gk enzyme that did not bind Pins; the alternative Anc-$GK1_{PID}$ bound Pins, as did the ML sequence. These results indicate that the inferred trajectory of functional evolution and the phylogenetic interval during which protein scaffolding activity first evolved are both robust to statistical uncertainty about the precise ancestral sequences (Figure 7D-E).

Figure 6 (next page). The mitotic spindle-orienting GK protein interaction domain evolved from guanylate kinase (gk) enzymes. **A)** The GK_{PID} of the protein Discs-large (Dlg, blue) serves as a scaffold for spindle orientation by physically linking the localized cortical protein Pins (green) to astral microtubules (red) via the motor protein KHC-73 (black). **B)** Reduced phylogeny of the protein family containing gk enzymes (brown) and protein-binding GK_{PID}s (blue). Parentheses show the number of sequences in each clade. Reconstructed proteins Anc-gk_{dup} (the pre-duplication ancestor of gk enzymes and GK_{PID}s in animals/choanoflagellates), Anc-$GK1_{PID}$ and Anc-$GK2_{PID}$ (the GK_{PID} in the common ancestor of animals and choanoflagellates, and of animals, respectively) are marked as circles with approximate likelihood ratio support. Scale bar indicates number of substitutions per site. For unreduced phylogeny, see Figure 5. **C)** Anc-gk_{dup} (circles) is an active nucleotide kinase in a coupled enzyme assay for the reaction shown; Anc-$GK1_{PID}$ (boxes) is inactive. Activity of the human gk enzyme (triangles) is shown for reference. Error bars shown are SEM for three replicates. **D)** The more recent ancestral protein Anc-$GK1_{PID}$ (triangles) binds a 20 amino-acid peptide (see box) from the Pins protein in a fluorescent anisotropy assay, but Anc-gk_{dup} (boxes) does not. Pins binding by the GK_{PID} of the D. melanogaster Dlg protein is shown for reference. Error bars are SEM for three replicates. **E-H)** Evolution of spindle orientation function as assayed in cultured S2 cells that do not express endogenous Dlg protein. Cells were transfected with a putative spindle-orienting protein and scored for the alignment of the mitotic spindle (red, tubulin, visualized immunocytochemically) relative to the Pins cortical crescent (green, a GFP-tagged Pins-Ecd fusion). In the images (left), two cells are shown, the bottom one of which is dividing. The angle of the mitotic spindle (from 0°, precisely aligned, to 90°) was recorded in many dividing cells; the radial histogram (right) shows the distribution of observed angles among all cells scored with a given genotype. Cells with no transfected GK domain (**E**, negative control) or with anc-gk_{dup} (**G**) do not mediate spindle orientation. Cells transfected with the GK_{PID} of D. melanogaster Dlg (**F**, positive control) or with Anc-$GK1_{PID}$ (**H**) display robust spindle orientation.

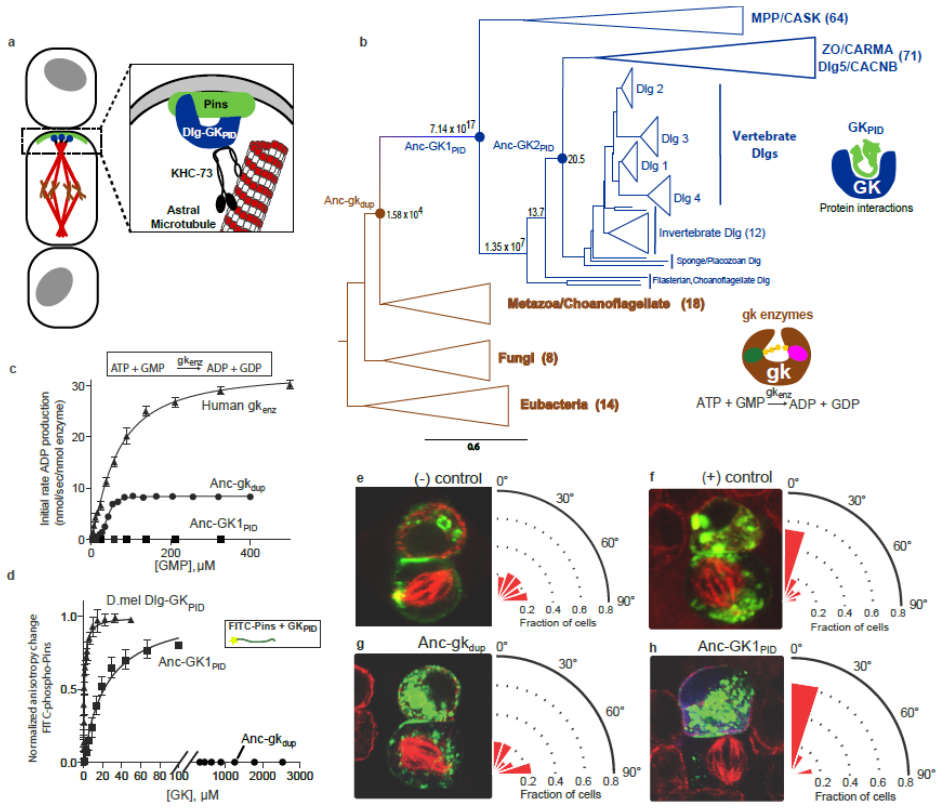
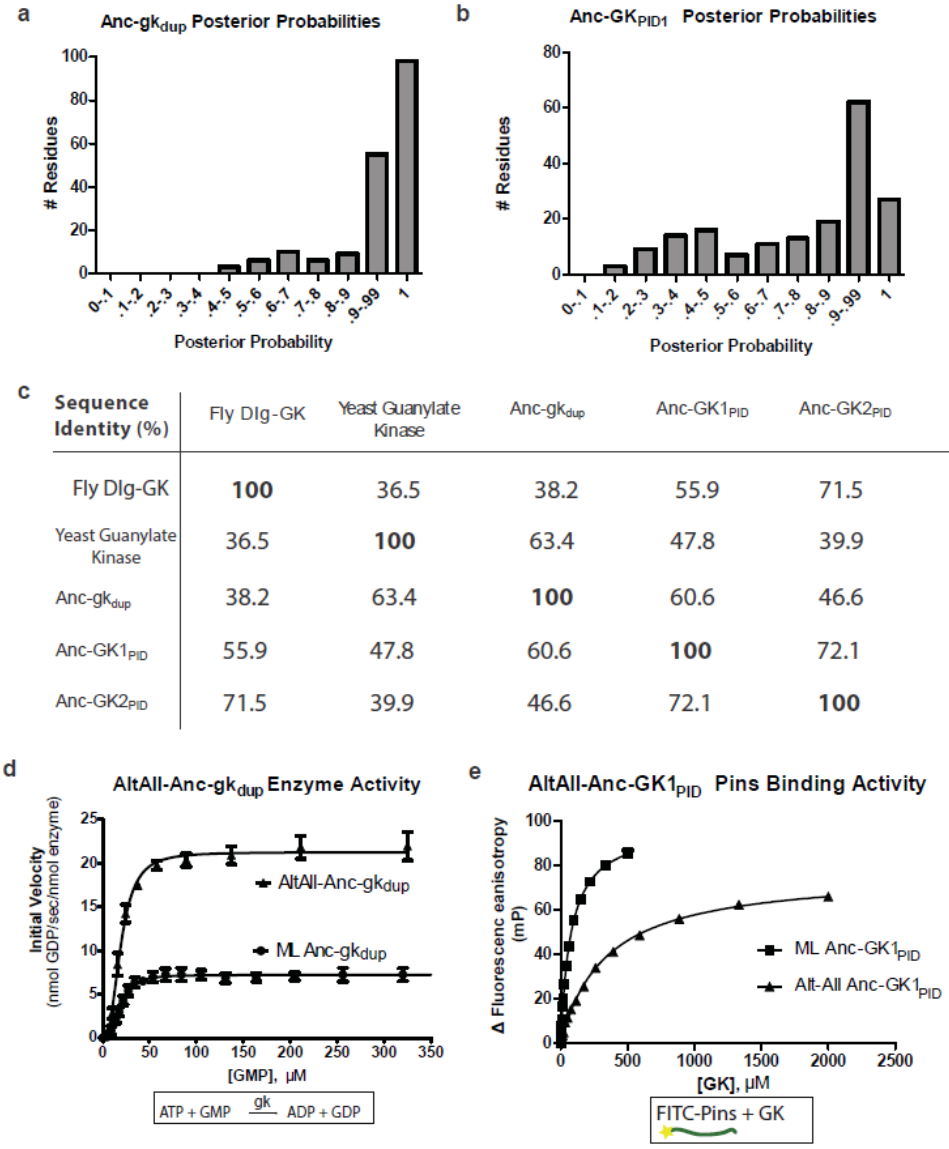


Figure 7 (next page). Characteristics and robustness of maximum likelihood reconstructions of Anc-gk$_{dup}$ and Anc-GK1$_{PID}$. A,B) The histogram shows the distribution over sites of posterior probability support for maximum likelihood amino acid states (see Appendix Tables A1 and A2 for full sequences and support). C) Sequence similarity of ancestral sequences to extant gk enzymes and GK$_{PID}$s. The table shows the percent of identical residues between each pair of sequences. D,E) Robustness of functional inferences about ancestral proteins to uncertainty about the sequence reconstruction. For both Anc-gk$_{dup}$ and Anc-GK1PID, alternate reconstructions (Alt-All) were synthesized which contains the next-best amino acid state at all sites with multiple plausible states, defined as having PP>0.2. The enzyme activity of Anc-gk$_{dup}$ in a coupled enzymatic assay for cofactor turnover (panel D) and the Pins-binding activity of Anc-GK1PID (panel E) in a fluorescence anisotropy assay are shown for both maximum likelihood and Alt-All reconstructions. Affinities and maximal velocities differ quantitatively, but the presence/absence of each property is robust to incorporation of uncertainty about the ancestral sequence.



Sequence Identity (%)	Fly Dlg-GK	Yeast Guanylate Kinase	Anc-gk$_{dup}$	Anc-GK1$_{PID}$	Anc-GK2$_{PID}$
Fly Dlg-GK	**100**	36.5	38.2	55.9	71.5
Yeast Guanylate Kinase	36.5	**100**	63.4	47.8	39.9
Anc-gk$_{dup}$	38.2	63.4	**100**	60.6	46.6
Anc-GK1$_{PID}$	55.9	47.8	60.6	**100**	72.1
Anc-GK2$_{PID}$	71.5	39.9	46.6	72.1	**100**

Taken together, these findings indicate that the capacity of the GK$_{PID}$ to bind Pins and orient the mitotic spindle arose before the divergence of choanoflagellates and before the advent of metazoans. This result is consistent with the hypothesis that the emergence of the GK$_{PID}$'s derived functions was an important step in the evolution of animal complexity. To understand whether other components of the spindle orientation machinery were also present at this time, we searched the genome of *Salpingoeca rosetta*, a choanoflagellate that forms spatially organized circular colonies of cells (43).

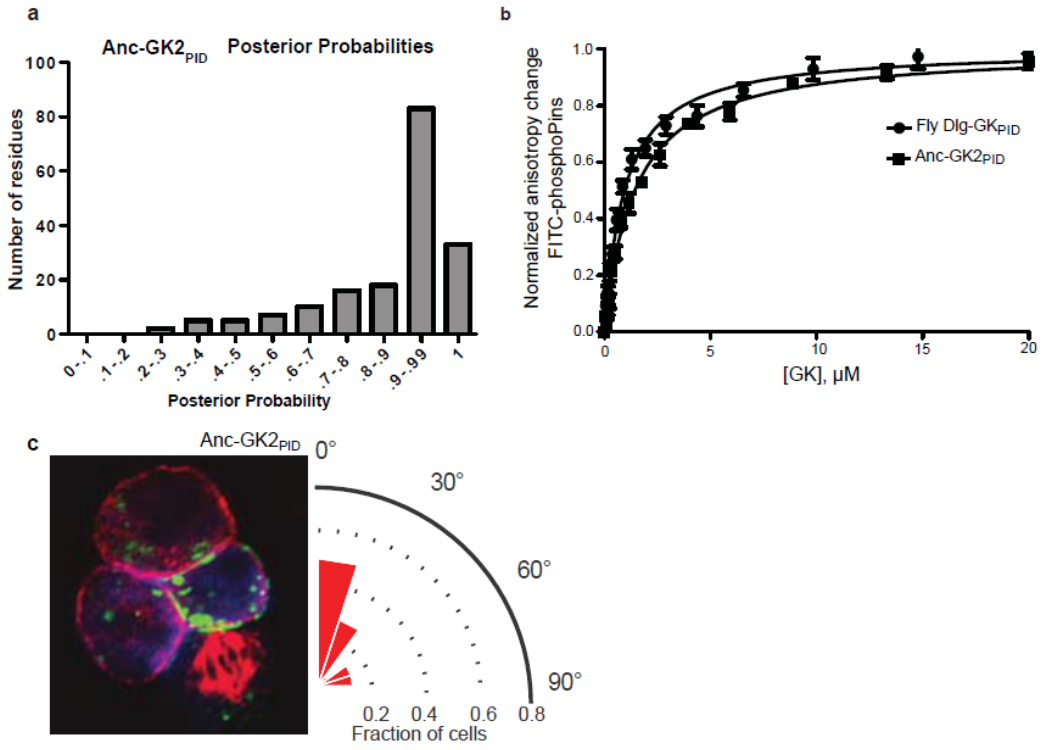
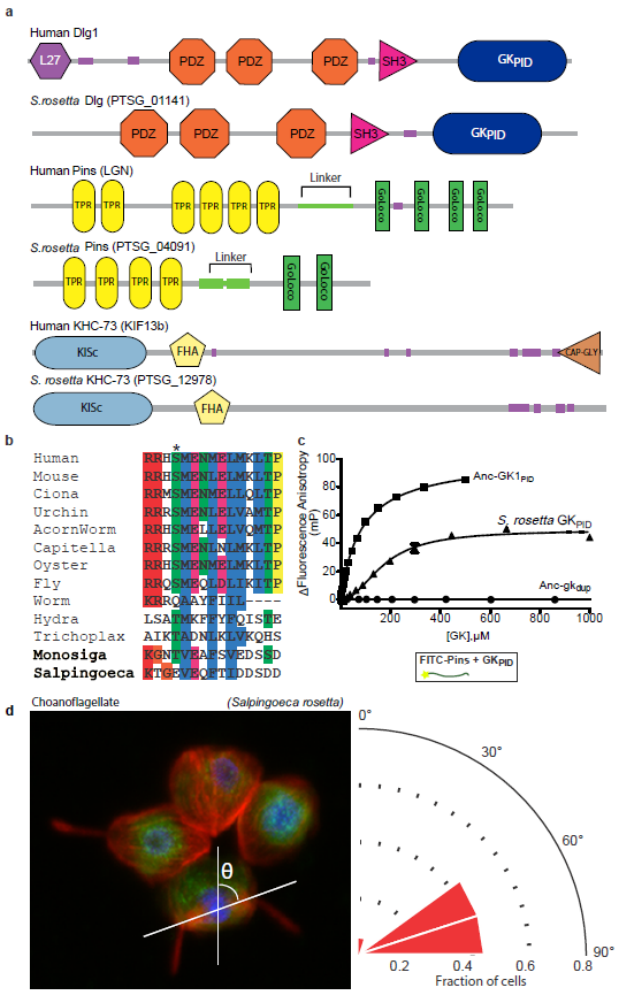


Figure 8. Properties of ancestral protein Anc-GK2$_{PID}$. A) Distribution of posterior probabilities for maximum likelihood amino acid inferences over sites. Average posterior probability = 0.87. B) Anc-GK2PID (squares) binds the Pins peptide ligand with high affinity in a fluorescence anisotropy assay. Binding by the D. melanogaster Dlg GKPID is shown for comparison (circles). Error bars are standard error of the mean of three replicates. C) Anc-GKdom2 is capable of orienting the mitotic spindle similar to extant controls (see Fig 5F).

We found that *S. rosetta* contains genes for Dlg, Pins and KHC-73-like proteins with similar sequences and domain structures (Fig. 9A). The Pins-like protein contains a recognizable but not strictly conserved version of the canonical Dlg binding motif found in Bilateria; the phosphorylated residue S436, which serves as a negatively charged anchor on Pins for binding to the Dlg GK$_{PID}$ in metazoans, is a negatively charged phosphomimic glutamic acid in *S. rosetta* (Fig. 9B). We expressed and characterized the GK$_{PID}$ of the S. rosetta Dlg-like protein in vitro and found that it is capable of binding metazoan Pins peptides with moderate affinity (Fig 9C), corroborating our observation with the

reconstructed Anc-GK1$_{PID}$ that Pins-binding capacity originated before the common ancestor of metazoa and choanoflagellates. The key genes whose products direct positioning of the mitotic spindle in extant animals are therefore older than the last common ancestor of animals and choanoflagellates. We also analyzed spindle orientation in dividing S. rosetta colonial cells and found that they orient their mitotic spindles in the plane of the colony (Fig. 9E) and in relation to the position of the flagellum. Taken together, these data suggest that mitotic spindle orientation itself is likely to have originated before the last common ancestor of animals and choanoflagellates.

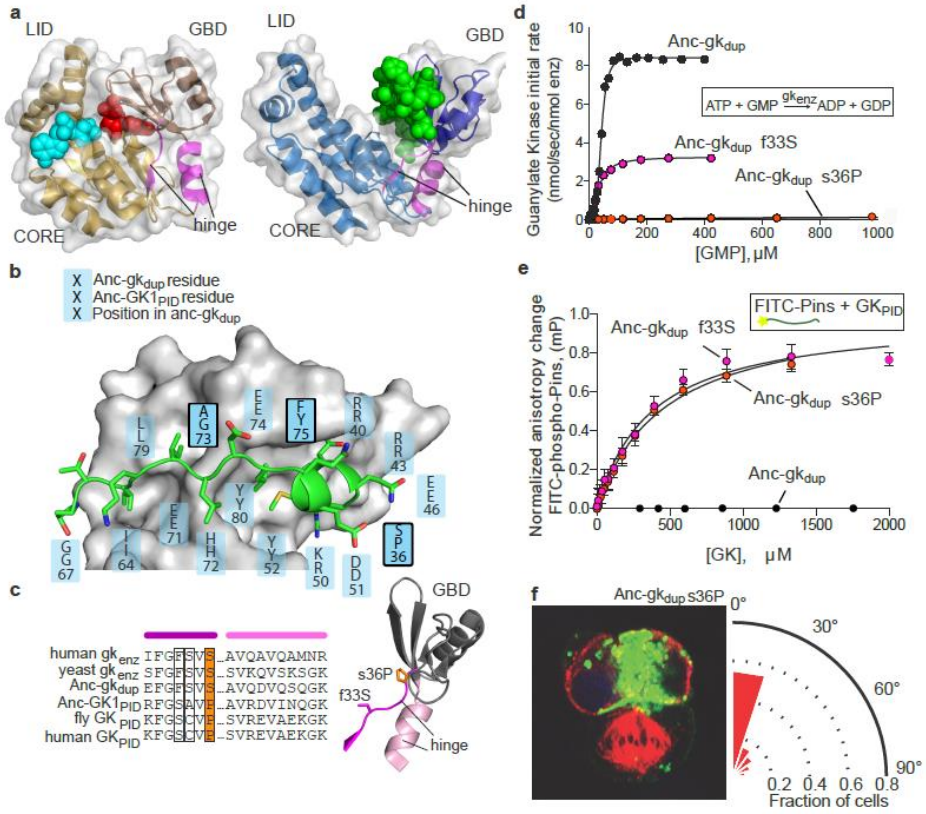
Figure 9 (next page). Choanoflagellates possess spindle orientation machinary. **A)** The choanoflagellate *Salpingoeca rosetta* genome encodes orthologs of human spindle orientation genes, including Dlg, Pins, and KHC-73. The domain architecture of each protein is shown, as inferred using the SMART database. Each pair of proteins is a reciprocal best-hit between the two species' genomes using BLAST. Small purple rectangles, low complexity regions. **B)** Aligned sequences from the linker portion of Pins (see panel A), which binds to Dlg. Colors highlight identical or biochemically conservative residues. Asterisk, phosphorylated or negatively charged residue 436, which anchors Dlg binding. For complete species names and accessions, see Supplementary Information Table 3. **C)** The purified GK$_{PID}$ from the choanoflagellate Dlg ortholog (squares) binds fluorescently labeled D. melanogaster Pins. Binding by Anc-GK1$_{PID}$ (circles) and lack of binding by Anc-gk$_{dup}$ (triangles) are shown for comparison. Error bars, SEM for three replicates. **D)** Cells from the colonial choanoflagellate S. rosetta orient their mitotic spindles. The image (top) shows one representative colony; the cell at bottom center is mitotic, as evidenced by condensed DNA (blue, DAPI) without a defined nuclear envelope (green, visualized using anti-nuclear pore complex). Red, mitotic spindle visualized using anti-tubulin. The angle of the mitotic spindle (solid white line) was measured relative to a line perpendicular to the plane of the colony extending through the colony's center (dashed line). The histogram shows the distribution of spindle angles among all dividing S. rosetta cells measured, with 90° representing perfect alignment relative to the colony ring.



We next sought to identify the genetic causes for the shift in function from enzyme to spindle-orienting scaffold protein. Seventy-one amino acid substitutions occurred during the interval between Anc-gk_{dup} and Anc-$GK1_{PID}$. To identify functionally causal substitutions, we relied on both structure-function information and the phylogenetic pattern of sequence conservation and divergence. Extant gk enzymes contain two nucleotide binding lobes connected by a flexible hinge region around a central catalytic core (Fig. 10A). In crystal structures of gk enzyme in the absence of nucleotide substrate, the binding lobes are separated from each other in an open conformation (44). Upon nucleotide binding, the lobes move inwards and occupy a closed conformation, bringing GMP and its

co-substrate ATP together and allowing catalysis to occur (45). In contrast, the GK_{PID} remains constitutively in the open conformation, and Pins – which is considerably larger than the enzyme's nucleotide ligands – binds to the exposed surface of the guanylate-binding lobe (12). (Fig. 10A) We reasoned that the substitutions that caused the functional transition from enzyme to scaffold might have affected residues that constitute the guanylate/Pins binding interface itself or, alternatively, residues in the hinge that determine the orientation of the lobes relative to each other and therefore affect the size and geometry of the cleft in which the ligands bind.

Figure 10 (next page). One historical substitution in the hinge is sufficient to convert a guanylate kinase to a spindle-orienting GK_{PID}. **A)** The common structural architecture of all gk/GK_{PID} family members consists of a catalytic core, two binding lobes (the GMP-binding domain, GBD, shown in dark hue, and the ATP-binding lid), and a flexible hinge region, which connects the GBD to the core and comprises two segments of contiguous residues (magenta). Left: in gk enzymes bound to GMP (red spheres), the lobes adopt a closed conformation, bringing GMP and ATP (cyan spheres) adjacent to each other in the core. Right: the GK_{PID} has an open conformation; Pins (green spheres) binds to the surface of the GBD in the open cleft between the two binding lobes. Structures shown are mouse gk enzyme (brown, PDB 1LVG) and the GK_{PID} from rat Dlg1 (blue, 3UAT). **B)** Most residues in Anc-$GK1_{PID}$ that bind Pins (blue boxes) are unchanged from the homologous residues in Anc-gk_{dup}. White surface, d. melanogaster Dlg-GK_{PID} surface (3TVT). Green, Pins peptide. Residues with historical substitutions between the two ancestral proteins are outlined. **C)** In the hinge region, two historical substitutions (outlined and colored) were conserved in the ancestral state in extant enzymes and a different state in extant GK_{PID}s. Colored bars above the sequence indicate position in the protein structure (right). Hinge segments are shown in pink and the GMP-binding lobe in gray. **D,E)** Introducing historical substitution s36P into Anc-gk_{dup} abolishes guanylate kinase activity (D) and confers binding to fluorescently labeled Pins **(E)**. Nearby substitution f33S confers binding but does not abolish enzyme activity. Error bars are SEM of 3 replicates. Substitutions in the binding interface do not recapitulate the evolution of Pins binding and the loss of gk activity (Fig. 11). **F)**. Introducing s36P into Anc-gk_{dup} confers full capacity to drive orientation of the mitotic spindle (compare Fig. 6).



In these regions of the protein, only five amino acid changes occurred between Anc-gk$_{dup}$ and Anc-GK1$_{PID}$ that are conserved among descendant GK$_{PID}$s (Fig. 10B, C). To test these substitutions' functional importance, we introduced the derived states individually into Anc-gk$_{dup}$ and characterized their effects on guanylate kinase activity, Pins binding, and spindle orientation. We found that substitution s36P was sufficient to nearly abolish the catalytic activity of Anc-gk$_{dup}$ and to establish moderate-affinity Pins binding (Figs. 10D,E; lower and upper case residue symbols denote ancestral and derived states, respectively). Remarkably, this single substitution was also sufficient to confer on Anc-gk$_{dup}$ the capacity to robustly mediate spindle orientation in cultured cells (Fig. 10F).

Substitution f33S, also in the hinge, had similar but weaker functional effects, conferring Pins binding and decreasing but not abolishing enzyme activity; combining f33S with s36P, however, did not further recapitulate the change in function beyond that caused by s36P alone (Fig. 10D,E; Fig 11A). Substitutions s34C, a73G and f75Y caused minor reductions in enzyme activity but did not confer strong Pins-binding. (Figure 11B-E). These data indicate that the single historical substitution s36P is sufficient to recapitulate the major functional transitions that occurred during the evolution of the GK_{PID} – loss of guanylate kinase activity, gain of a substantial portion of the Pins affinity observed in extant Dlg proteins, and acquisition of the capacity to orient the mitotic spindle in extant animal cells.

Discussion

How could a single amino acid change have caused this entirely new function to emerge? Substitution s36P does not appear to have constructed a new binding site for Pins, because the residues that compose the Pins-binding surface are almost entirely conserved from Anc-gk_{dup} to Anc-$GK1_{PID}$ (Figs. 10B), and the structure of the binding lobe is almost identical between the crystal structures of extant GK_{PID}s, gk apo-enzymes, and a gk apo-enzyme containing mutation S36P (Fig. 13B, ref. 8). Indeed, the surface that the GK_{PID} uses to bind Pins appears to represent a fortuitous redeployment of the gk enzyme's binding site for GMP, along with some adjacent surface (Fig. 13A, ref.45). The enzyme anchors GMP's phosphate group with hydrogen bonds to four clustered residues that form a positively charged pocket (R40, R43, Y80, Y82); in the GK_{PID}, these residues are conserved and form hydrogen bonds with the phosphate group of Pins' phospho-serine 436.

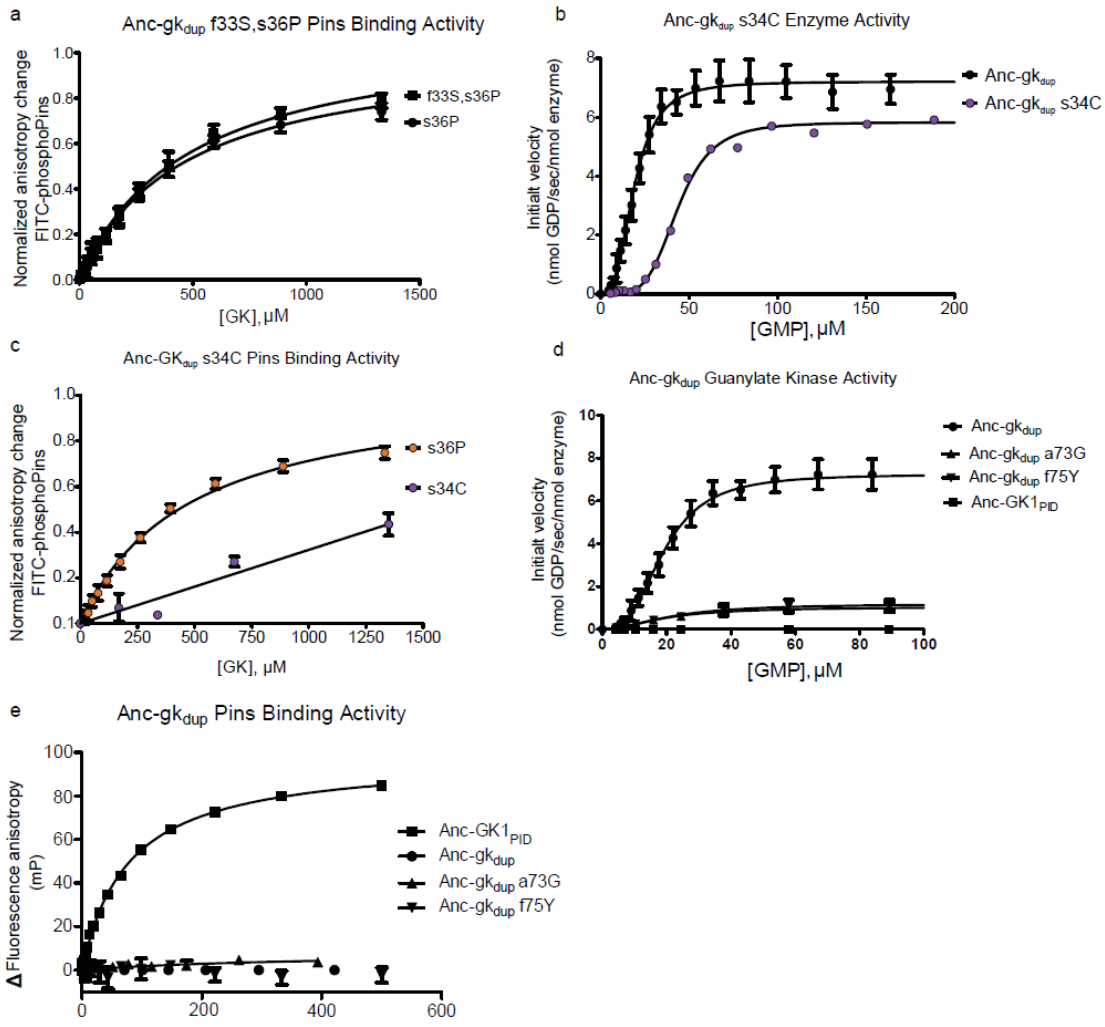


Figure 11. Introducing additional historical substitutions into AncGK$_{dup}$ does not confer GK$_{PID}$-like function. A) Combining hinge substitutions f33S with s36P in Anc-gk$_{dup}$ (squares) does not further improve Pins affinity compared to Anc-gk$_{dup}$-s36P in a fluorescence anisotropy assay with labeled Pins peptide ligand. B) The hinge substitution s34C (purple) does not abolish guanylate kinase activity in a coupled enzyme activity assay; Anc-gk$_{dup}$ is shown for comparison. C) s34C (purple) does not confer substantial Pins-binding affinity on Anc-gk$_{dup}$. Pins-binding activity relative to s36P (orange) in a fluorescence anisotropy assay. D,E). Anc-GK1PID Pins interface substitutions a73G and f75Y reduce gunaylate kinase enzyme activity (D) but do not confer Pins binding (E) when introduced into Anc-gk$_{dup}$. Error bars are standard error of the mean with three replicates.

At the other end of GMP, a hydrophobic groove in the enzyme makes van der Waals contacts to GMP's guanosine hydrophobic ring, and the same groove on the GK$_{PID}$ binds to a series of hydrophobic side chains on Pins. The GK$_{PID}$ also makes further hydrophobic

contacts to Pins in the groove beyond the GMP binding site, and there are just two additional hydrogen bonds, both from backbone atoms on Pins to polar residues on the GK_{PID} surface that were solvent-exposed in the enzyme. Thus, the binding surface for both GMP and Pins has a relatively simple structure and surface properties, with specificity established largely by the positively charged pocket on one end and the size and position of the hydrophobic groove; Pins shares with GMP the crucial surface properties to be accommodated by the same interface on the protein.

The binding surface for Pins therefore appears to have been present in latent form in the Anc-gk_{dup} enzyme. How then did s36P confer Pins binding? Several lines of evidence suggest that this substitution is likely to have altered the protein's dynamics and/or increased the relative occupancy of a conformation in which the latent binding site is exposed for peptide binding. First, residue 36 is located where the hinge joins the GMP-binding lobe (Figs. 10B, C, Fig 12A). This hinge is known to mediate the dynamic opening/closing of the binding lobes relative to each other (ref. 44, Fig. 10A). Second, the degree to which the domain is open or closed appears to be essential for function. In gk enzymes, closing is critical for catalysis, because it brings the nucleotide substrates close together in the protected active site 30; in GK_{PID}s, however, open conformations appear to be required for Pins binding, because the Pins peptide is significantly larger and is therefore predicted to sterically clash with the GK_{PID} when its lobes are close together.

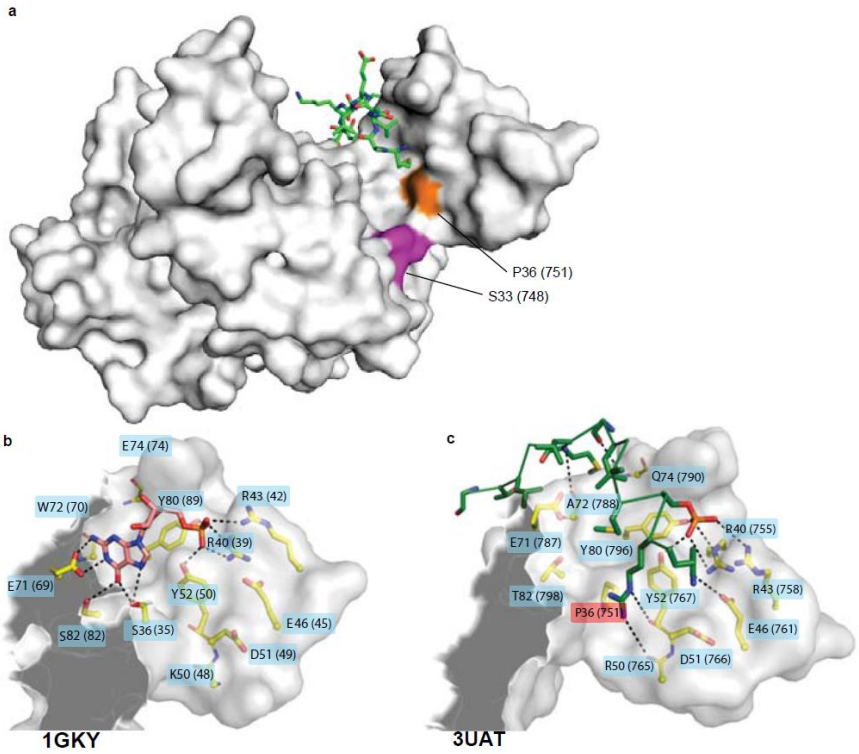


Figure 12. Structural context of key historical mutations. A) Location of historical hinge substitutions s36P and f33S. A. Residues Pro36 (orange) and Ser33 (purple) are shown on the crystal structure of the GKPID from rat Dlg (3UAT, white surface). Pins peptide ligand is shown in green. B, C). Similarity of GMP binding site in extant guanylate kinase enzyme to Pins binding site in extant GKPID. The guanylate binding domains (GBDs) of yeast gk enzyme (PDB 1GKY, panel B) and of rat Dlg GKPID (PDB 3UAT, panel C) is shown as white surface, with all side chains that contact either GMP or Pins as yellow sticks. Pink sticks show GMP; green ribbon shows Pins backbone, with the side chains of all Pins residues that contact the GKPID protein shown as sticks. The phosphate group on GMP and on Pins residue S436 are indicated. Black dotted lines, protein-ligand hydrogen bonds. Key substitution s36P is highlighted in pink.

Thirdly, substituting a proline would restrict backbone dihedral angles in the hinge, possibly altering the dynamics of the hinge and/or the distribution of conformations it occupies. Introducing a proline at residue 36 into extant gk enzymes has been shown to impede the GMP-induced closing motion, abolish enzyme activity, and to confer Pins binding (22). Because the effects of mutation s36P on the function of the ancestral gk enzyme are nearly identical to those it has on the extant enzyme, it is likely that similar

biophysical mechanisms pertain in the two proteins. Finally, mutations in extant and ancestral gk enzymes at other hinge residues 32 and 33, which make no direct contacts to the ligand, also reduce enzyme activity and confer Pins binding and/or spindle orientation, and the mutation at residue 32 has been shown to impede conformational closing, supporting a causal link between these phenomena (Figs. 10D-E, ref. 22).

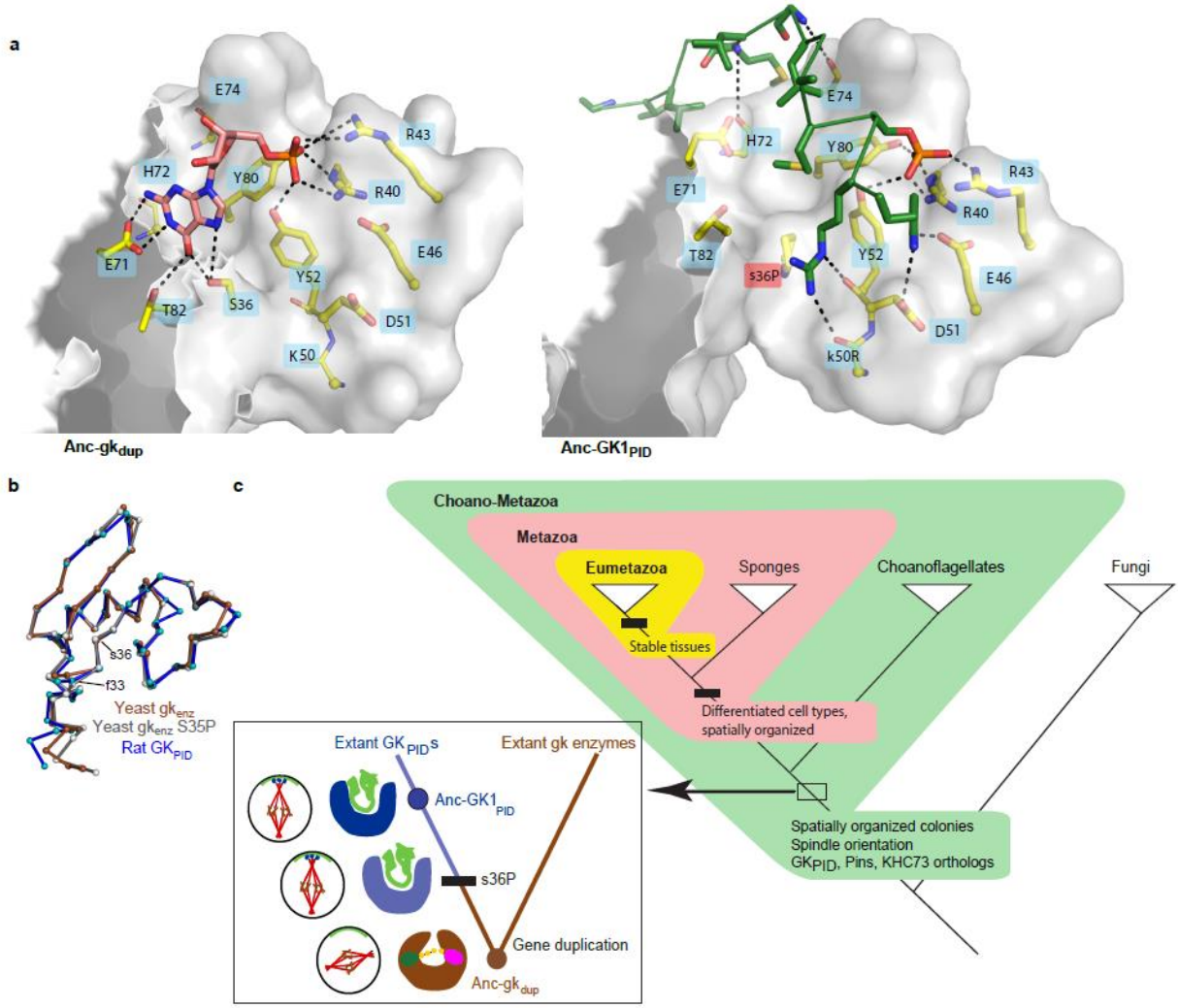
We therefore propose that substitution s36P conferred the novel scaffolding function of the GK_{PID} by restricting the ancestral enzyme's dynamic hinge motions and/or changing its occupancy of conformations with high affinity for Pins (Fig. 13D). By impeding GMP-induced closing, this substitution would have inhibited enzyme activity. Further, by changing the protein's dynamics or increasing the probability that it would occupy open conformations in which the latent peptide-binding interface is revealed, s36P would have conferred moderate affinity for Pins and spindle-orientation capacity, just as we observed. This scenario is consistent with recent findings that changes in conformational occupancy may play an important role in the evolution of protein function (46,47). Subsequent substitutions apparently fine-tuned Pins-binding affinity, yielding the higher-affinity Anc-$GK1_{PID}$ and Anc-$GK2_{PID}$. In addition to its likely effects on conformational occupancy and/or dynamics, Pro36 makes van der Waals contact to the Pins ligand (Fig. 12A), so it is possible that s36P may also have contributed to optimizing the latent Pins binding interface itself.

Our findings shed light on the timing and mechanisms by which the proteins that control animal spindle orientation evolved, and they have implications for of the evolution of organized multicellularity more generally. Our experiments establish that Anc-$GK1_{PID}$ – which predates the ancestor of choanoflagellates and animals – could bind Pins and

mediate spindle orientation in animal cells, implying that it also had the capacity to bind other components of the spindle orientation machinery that are anchored by Dlg

---

Figure 13 (next page). Evolution of GK_{PID}'s new function by unveiling a latent protein-binding site. **A)** The binding surface for Pins in GK_{PID}s is derived from the GMP-binding surface of gk enzymes. Homology models of Anc-gk_{dup} (left) and Anc-$GK1_{PID}$ (right) are shown as white surface, with all side chains that contact either GMP or Pins as yellow sticks. Pink sticks show GMP; green ribbon shows Pins backbone, with the side chains of all Pins residues that contact the GK protein shown as sticks. The phosphate group on GMP and on Pins residue S436 are indicated. Black dotted lines, protein-ligand hydrogen bonds. Key substitution s36P is highlighted in pink. The binding mode of extant enzymes and domains supports the same conclusion (see Figure 12). **B)** The structure of the hinge and GMP/Pins-binding lobes is conserved between the Pins-bound GK_{PID} (blue, rat Dlg, 3UAT), the apo-gk enzyme (brown, S. cerevisiae guanylate kinase 1EX6), and the apo-gk-s36P mutant (gray, 4F4J), all in the open conformation. **C)** Summary of the evolution of GK_{PID}'s spindle orientation capability. Major clades on the phylogeny of metazoans and closely related taxa are indicated; small rectangles mark the origin of shared derived characters related to organized multicellularity. Inset, evolution of the GK_{PID} (blue) after $_{dup}$lication of an ancestral gk enzyme (brown) during the interval marked by the open rectangle. Substitution s36P was sufficient to abolish enzyme activity, confer binding to Pins peptide (green), and confer the capacity to mediate spindle orientation (schematic of cell at left).



in extant metazoans. These results indicate that the capacity of the GK_{PID} to carry out the molecular functions related to spindle orientation existed before the advent of organized multicellularity. We observed that animals and choanoflagellates both display some form of spindle orientation; although the mechanism in the latter taxon is unknown, orthologs of Dlg, Pins and KHC-73 are present in both taxa, suggesting that spindle orientation itself also predates the choanoflagellate/metazoan ancestor and the evolution of organized multicellularity per se. These genes all appeared by duplication and divergence during the long interval between the ancient last common ancestor of Fungi and Metazoa and the choanoflagellate/metazoan ancestor, so the precise order in which they evolved cannot be

resolved by current data. It is therefore unclear whether the single s36P mutation served as a "final trigger" in the evolution of spindle orientation or an intermediate event that conferred on Anc-GK1$_{PID}$ the capacity to serve as a scaffold for partners that had not yet evolved their spindle-orienting forms. In either case, our results indicate that a very simple genetic change was sufficient to cause a key molecular prerequisite for organized multicellularity to evolve. Such a radical evolutionary shift was possible because the mutation had only to reveal a previously hidden surface that had the fortuitous capacity to mediate molecular interactions now indispensable for the complexity of all multicellular animals.

Methods

Phylogenetics, ancestral protein reconstruction, expression, and homology modeling. Annotated protein sequences for 224 guanylate kinases and GK$_{PID}$s were downloaded from UniPROTKB/TrEMBL, GenBank, the JGI genome browser, and Ensemble databases. Amino acid sequences were aligned using MUSCLE (51) followed by manual curation and removal of lineage-specific indels. For species and accessions used, see Supplementary Table S3. Guanylate kinase sequences were trimmed to include only the active gk domain predicted by the Simple Modular Architecture Research Tool (SMART) (52). The phylogeny was inferred by maximum likelihood using PhyML v2.4.5 (53) and the WAG model with gamma-distributed rate variation and empirical state frequencies, which was selected using ProtTest software and the AIC criterion. Statistical support for each node was evaluated by obtaining the approximate likelihood ratio (the likelihood of the best tree with the node divided by the likelihood of the best tree without the node) and the chi-squared confidence statistic derived from that ratio (54). Ancestral protein sequences and

their posterior probability distributions were inferred by the maximum likelihood/empirical Bayes method (55), assuming the ML phylogeny and the best-fit model, using PAML v3.13 and Lazarus software (56) . Average probabilities were calculated across all GK sites except those containing indels. Plausible alternative non-ML states were defined as those with posterior probability >0.20. Alternate ancestral sequences at each node contained all plausible alternative non-ML residues in a single protein.

Coding sequences for ancestral proteins were inferred assuming *E. coli* codon bias, synthesized (Genscript), and inserted into pBH plasmid vector with a hexa-His tag for E. coli expression at 20 °C. Protein purification was carried out using sequential NiNTA affinity, anion exchange, and size-exclusion chromatographies. All proteins eluted as predicted monomers from the size-exclusion column at purity >95% by Coomassie staining of an SDS-PAGE gel. Proteins were concentrated using Vivaspin concentrators (Sigma-Aldrich), flash frozen in liquid nitrogen, and stored at −80 °C in buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 1 mM DTT).

Homology modeling was performed with SWISS-MODEL structure homology tool, (http://swissmodel.expasy.org/). 1GKY was used as the template for Anc-gk_{dup} and 3UAT for Anc-$GK1_{PID}$.

Protein binding fluorescence anisotropy assay. Binding of *D. melanogaster* Pins was assayed by fluorescence anisotropy on a Tecan Sapphire plate reader equipped with automatic polarizers using default settings for anisotropy assays. A FITC-labeled peptide (GVRVRRQ(pS)MEQLDLIKITPD, Genscript) of the fly Pins-Linker peptide (0.25 μM) was incubated with increasing concentrations of GK protein in phospho-buffered saline

solution with 1mM DTT. A one-site binding model was used to fit the data and infer binding affinity in Graphpad Prism.

Spindle Orientation Assays. Maintenance of S2 cells, construction of expression plasmids, and cell-adhesion/spindle orientation assays were performed as detailed previously (17). S2 cells were transfected with constructs for FLAG- or HA-tagged GK protein and for Pins-GFP-Echinoid fusion protein, using Effectene reagent (Qiagen) with 0.4–1 μg total DNA for 24–48 h. Endogenous Dlg and Aurora A were knocked down using RNAi: transfected cells were incubated for 1 h in serum-free media containing approximately 1 μg RNAi followed by 72 h in normal growth media. Protein expression was then induced by adding 500 μM CuSO4 for 24 h. Cell-adhesion clustering was induced by constant rotation at approximately 175 rpm for 1–3 h. Pins fusion protein was visualized by fluorescence (excitation 488 nm, emission 509 nm). Mitotic spindles were visualized using rat anti-tubulin (Abcam 1:500) and goat anti-rat conjugated to Alexa 555 (Life Technologies, 1:500). In addition, GK was visualized using mouse anti-FLAG or anti-HA (Sigma 1:500) and chicken anti-mouse:Alexa 647 (Life Technologies 1:200); histones were visualized using rabbit antiphospho-histone3 (Upstate 1:8000) and donkey anti-rabbit:Alexa 488 (Life Technologies, 1:500). For immunostaining, clustered cells were fixed in 4% paraformaldehyde for 20 min, washed, and incubated with primary antibodies overnight at 4 °C in buffer (0.01% saponin plus 0.1% albumin diluted in phosphate-buffered saline). Slides were subsequently washed and fluorescently linked secondary antibodies were added for 2 h at room temperature. Finally, slides were again washed and mounted using Vectashield Hardset medium (Vector Laboratories). All images were collected using a Biorad Radiance 2100 confocal microscope with a 1.4NA 60X objective .

The spindle angles were measured for ~20 cells for each condition and displayed in a radial histogram.

Guanylate kinase activity assay. We used a coupled enzyme assay, as described previously (57), which quantifies release of ADP as the guanylate kinase-catalyzed reaction proceeds by coupling it to pyruvate kinase- and lactate dehydrogenase-catalyzed reactions and measuring the consequent oxidation of NADH by following the decrease in absorbance at 340 nm, measured on a Tecan Safire plate reader. Guanylate kinase enzyme was at 50-200 nM in assay buffer (100 mM Tris, pH 7.5, 100 mM KCl, 10 mM MgCl2, 1.5 mM sodium phosphoenolpyruvate, 300 mM NADH, 4 mM ATP, and 100 units pyruvate kinase and 100 units lactate dehydrogenase).Initial GMP concentrations ranged from 500 nM to 1 mM. The reaction was initiated by adding GMP and briefly mixing. Reactions were carried out at 30 °C and measured 30 times at 15s intervals. Data were analyzed and plotted using GraphPad Prism assuming Michaelis-Menten kinetics. Reaction rates are plotted as initial rate of ADP production.

Choanoflagellate culture, spindle orientation assays, and gene ortholog inference. Growth medium was prepared in artificial sea water and *S. rosetta* cultures (ATCC 50818) consisting primarily of chain colonies and slow swimmers were maintained by passaging 2 mL of culture into 18 mL fresh medium every day (58). Rosette colonies were produced by inoculating *S. rosetta* chain colonies with *Algoriphagus machipongonensis* bacteria[26]. Log phase *S. rosetta* cells were treated with 0.33uM Nocodazole (Sigma M1404) for 18 hours at RT. Cells were pelleted by centrifugation for 5 minutes at 2000 x g, and washed 3 times in artificial sea water to remove drug. Cells were allowed to recover for 30, 45, or 60 minutes at RT before fixation. Approximately 0.1 ml of cells were applied to poly-L-

lysine-coated 96-well plates and left to attach for 30 min. Cells were fixed for 5 min with .2 ml 6% acetone, and then for 20 min with .2 ml 4% formaldehyde. Acetone and formaldehyde were diluted in artificial seawater, pH 8.0. Wells were washed gently four times with 1 ml washing buffer (100 mM PIPES at pH 6.9, 1 mM EGTA, and 0.1 mM $MgSO_4$) and incubated for 30 min in 1 ml blocking buffer (washing buffer with 1% BSA, 0.3% Triton X-100). Cells were incubated with primary antibodies diluted in 0.15 ml blocking buffer for 1 h, washed four times with .2 ml of blocking buffer, and incubated for 1 hour in the dark with fluorescent secondary antibodies (1:1000 in blocking buffer, Alexa Fluor 488 goat antimouse, and Alexa Fluor 568 goat anti-rabbit; Invitrogen). Wells were washed three times with washing buffer, blocked with .2mL DAPI for 5 minutes, and washed twice more. The following primary antibodies were used: Mouse monoclonal antibody against β-tubulin (E7, 1:100; Developmental Studies Hybridoma Bank) and nuclear pore complexes (1:100, Covance). Images were taken with a 63× oil immersion objective on a Leica DMI6000 B inverted compound microscope and Leica DFC350 FX camera.

Choanoflagellates orthologs to metazoan guanylate kinase, Dlg, Pins and KHC-73 proteins were inferred by using the NCBI BLAST tool using human or drosophila protein sequences as the query. Candidate ortholog subjects were checked for domain architecture similarity to the query protein using SMART (52) and then verified with a reciprocal BLAST.

CHAPTER III

EVOLUTION OF DLG-GK DOMAIN PHOSPHO-SPECIFICITY

Introduction

Within the Dlg family lineage of the MAGUKs a defining phenotype of the GK domain is its ability to bind phosphorylated protein ligands (13). Although it is currently unknown if every Dlg family GK domain ligand is phosphorylated, current crystal structures and genetic evidence point towards the GK domain as a phospho-peptide binder which excludes the unphosphorylated ligand. Examinations of the physical nature of the GK domain ligand interactions would suggest that the negatively charged phosphate-binding pocket, derived from the guanylate kinase ancestor's guanosine binding pocket, are responsible for this specificity by providing the majority of the binding energy which is specific to the phosphate (12). Exclusion of this phosphate would exclude this binding energy would result in the loss of a significant portion of the binding energy resulting in functionally disparate binding affinities.

The phospho-specific behavior of the Dlg-GK domain has been best characterized in the GK / Pins interaction involved in spindle orientation (Figure 14). In the *Drosophila melanogaster* neuroblast model of spindle orientation, Pins is phosphorylated by the mitotic kinase Aurora A at residue S436, allowing the formation of the GK/Pins/KHC-73 pathway to form and initiate spindle orientation. Aurora A is only active during mitosis, allowing Dlg-Pins interaction only during mitosis, conferring cell-cycle specific regulation to the interaction. This phosphorylation event has been shown to be necessary not just for robust spindle orientation, but also for the overall development of the larvae. Flies deficient

in Aurora A or ablation of the Aurora A phosphorylation site on Pins results over-proliferation of neuroblasts during development and eventual death of the organism (17).

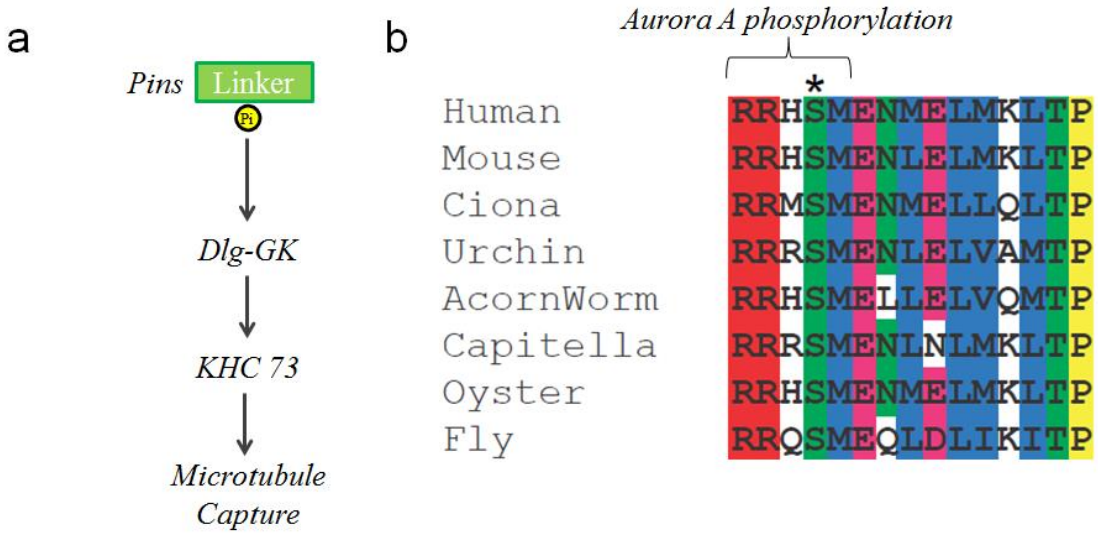


Figure 14. The Dlg-GK domain is specific to phosphorylated Pins. **A**. Phosphorylation of Pins at residue 436 is necessary for the Pins / GK spindle orientation pathway. Lack of phosphorylation at this residue ablates spindle orientation in cells. **B**. The GK domain binding site on Pins is strictly conserved in bilaterally symmetrical organisms with the Aurora A phosphorylation motif a critical portion of this binding sequence. Phosphorylated residue 436 is shown with an asterisk.

Further evidence for the importance of phosphorylation in the Pins / GK interaction can be seen at the sequence level at the GK binding site on Pins itself. Like many protein kinases, Aurora A possesses a phosphorylation consensus sequence located on the potential substrate which determines which substrates will be bound and phosphorylated. For Aurora A, this sequence is R-R/K-X-S/T-β (49) and these residues make up nearly half of the binding interaction between Pins and GK domain itself (Fig 14 B, ref. 13). Furthermore, this consensus site itself is strictly conserved throughout all of Bilateria highlighting its strict importance throughout most of metazoa. Experimental mutation of these residues,

most specifically the phosphorylatable serine , results in drastically lowered binding affinity, congruent with the whole-fly experiments discussed previously (17).

Although it is clear that the phosphorylation of Pins and subsequent phospho-specific binding by the GK is present and necessary by most of Metazoa, it remained unknown exactly when and how the GK domain evolved its specific phospho-ligand binding ability within the context of the results of the previous chapter. While the previous study elucidated the origins of Pins binding and spindle orientation capacity, it remained unclear as to when phospho-specificity was evolved in addition to general spindle orientation and Pins binding. To answer this question, we sought to assay the ancestors tested previously again to assay for not just the ability to bind phosphorylated-Pins, but also an unphosphorylated Pins in order to discover the evolutionary origin of GK domain phospho-specificity.

The initial characterization of general GK domain protein binding and spindle orientation from the previous chapter indicated that general Pins binding and spindle orientation capability first was present in the ancestor to all GK domains, Anc-GK1$_{PID}$, and that the affinity for Pins increased to near extant affinities with the more recent ancestor to all Dlg proteins, Anc-GK2$_{PID}$. Although we noted reasonably high affinity for phosphorylated Pins in Anc-GK1$_{PID}$ and Anc-GK2$_{PID}$, it remained possible that these proteins retained significant affinity for unphosphorylated Pins as well. We also noted that the conservation pattern of the Aurora A phosphorylation motif on Pins, which makes up a large portion of the GK binding site, is conserved most strongly in Bilateria (Figure 14B, suggesting that phospho-specific GK binding may have evolved much later than the earliest GK domain, Anc-GK1$_{PID}$, which predates the choanoflagellate/ Metazoa divergence.

Phospho-specificity Characterization of Ancestral GK Proteins

To determine when phospho-specific Pins binding evolved in the GK domain, we screened the major ancestral GK proteins starting from Anc-gk_{dup} onward for the ability to bind unphosphorylated Pins in the in vitro anisotropy assay as described previously and testing for mitotic spindle orientation capability in S2 cells either treated with RNAi against Aurora A or transfecting Pins construct in which the phosphorylated serine is ablated to unphosphorylatable alanine (S436A). Given that Anc-gk_{dup} did not possess binding activity or orient the mitotic spindle, we predicted that Anc-gk_{dup} would also not respond to unphosphorylated Pins and would also not orient the mitotic spindle. As expected, I found that Anc-gk_{dup} did not orient the spindle with unphosphorylated Pins (Figure 15A). I next tested derived ancestors AncGK1$_{PID}$ and AncGK2$_{PID}$ for phospho-specificity and found, somewhat surprisingly, that AncGK1$_{PID}$ did not bind unphosphorylated Pins with any detectable affinity and did not orient the mitotic spindle in the absence of phosphorylation while robustly orienting the mitotic spindle in the presence of phosphorylated Pins(Fig 15B-C). The further derived ancestor Anc-GK2$_{PID}$ also did not orient the mitotic spindle without Pins phosphorylation, and bound unphosphorylated Pins with ~400-fold less affinity than phospho-Pins (Figure 15C). Taken together, these results suggest that the ability to bind exclude unphosphorylated Pins interactions evolved concomitantly with Pins binding in the divergence from Anc-gk_{dup} to Anc-GK1$_{PID}$.

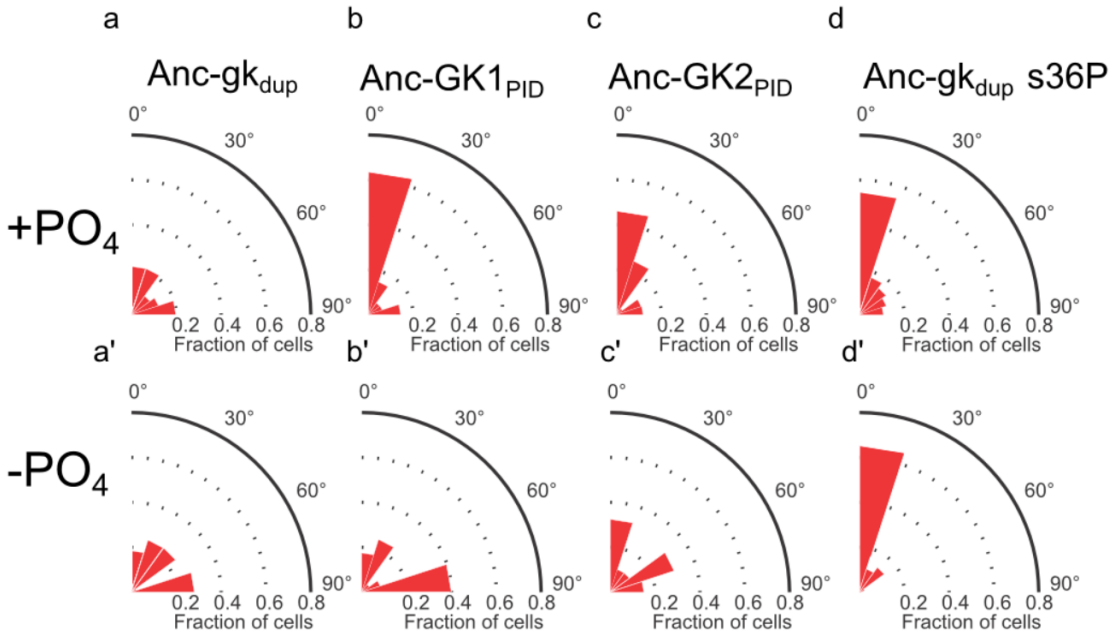


Figure 15. Phospho-specificity characterization of GK ancestors by spindle orientation in cultured *Drosophila* S2 cells. Top row are conditions in which Pins is phosphorylated normally by Aurora A kinase, while the bottom row represent conditions in which the phosphorylatable S436 was removed. **(A-A')** The $_{dup}$licated guanylate kinase ancestor possess no spindle orientation regardless of Pins phosphorylation state. Derived GK domain ancestors Anc-GK1$_{PID}$ **(B,B')** and Anc-GK2$_{PID}$ **(C,C')** orient the spindle only in the presence Pins phosphorylation.(**D-D'**) Addition of the hinge s36P substitution to the Anc-gk$_{dup}$ confers spindle orientation activity regardless of Pins phosphorylation.

Identification of Sufficient Anc-GK1$_{PID}$ Phospho-specificity Residues

I next sought to determine the genetic determinants of phospho-specificity in Anc-GK1$_{PID}$. With the results from the previous chapter II in mind, I reasoned the historical substitution s36P may also confer the ability to discriminate between Pins phospho-states in addition to conferring general Pins binding by utilizing the pre-built phosphate binding pocket from Anc-gk$_{dup}$'s guanosine binding domain. Sequence analysis as well as homology models of Anc-gk$_{dup}$ indicate a potential in-built phosphorylation recognition ability by utilizing the residues previously utilized to bind the phosphate of GMP. As the experiments with s36P appear to expose this latent binding site, I reasoned that Anc-gk$_{dup}$ s36P may also have a level of phospho-specificity. To investigate this, I next assayed Anc-

gk_{dup} s36P for spindle orientation and Pins binding without phosphorylation. Surprisingly, we found that Anc-gk_{dup} s36P both oriented the mitotic spindle and bound unphosphorylated Pins (Figure 15D). Interestingly, Anc-gk_{dup} s36P had nearly the same affinity as phosphorylated Pins suggesting that there was no level of intrinsic phospho-specificity granted by s36P (Figure 18). This result was rather unexpected in light of the results from chapter II, given that s36P appears to expose a Pins binding site very similar observed in extant Pins-GK interactions. From this observation, I reasoned that other substitutions must be necessary in order to fully exploit the full Pins binding site. Because the Pins-GK interaction takes place on the GBD subdomain , I reasoned that substitutions involved in Pins phospho-specificity would be located within the GBD and are possibly residues that directly contact Pins. Substitutions which met this criteria had been previously screened for conferral of Pins binding to Anc-gk_{dup} and found to have no effect (see figure 10B, chapter II). Because these individual substitutions had no individual effect, I reasoned that these conserved mutations may be "building" upon the activity granted by s36P and may be involved in a fine-tuning mechanism upon Pins-binding. With this logic I selected 7 diagnostic substitutions from the GBD to be screened for the conferral of phospho-specificity upon the more general binding of Anc-gk_{dup} s36P in the context of mitotic spindle orientation. Because many of these substitutions had been tested previously and found to have no effect individually, I also reasoned that it was likely there was a group of substitutions necessary rather than any single one substitution. With this in mind, I screened for phospho-specificity starting from the entire group of substitutions rather than single substitutions.

I found that the seven candidate substitutions when added to s36P conferred a moderate level of phospho-specificity (Figure 16a) that did not orient the mitotic spindle as strongly as controls in the phosphorylated condition, but was able to exclude the unphosphorylated condition like that of extant controls. From this initial result, combinatorial screening was performed to isolate the minimal number of substitutions which would orient the spindle in the phosphorylated condition, but would not orient in the absence of phosphorylation. Interestingly, among the initial group of 7 substitutions, I found that two specific subgroups of substitutions appeared to have discrete effects, substitutions which completely ablated spindle orientation in both conditions (Group 1, Fig.16B) and substitutions which had no effect on phospho-specific spindle orientation (i.e. oriented the spindle in both phosphorylated and unphosphorylated conditions, Fig. 16C). Because the group 1 substitutions completely ablated all spindle orientation, we hypothesized that these substitutions may be excluding the unphosphorylated condition when in conjunction with other substitutions from the original group of seven. Continued screening found that three substitutions, k41P, a73G, and f75Y, from these groups was sufficient to recapitulate complete phospho-specific spindle orientation like that of extant controls and more robust than that of the original group of 7 candidate substitutions (Figure 17A). Furthermore, we found that these substitutions had no such effect without s36P; reversion of the hinge proline back to serine resulted in a loss of all spindle orientation activity regardless of phosphorylation state, suggesting that the substitutions modify the activity granted by s36P itself (Fig 17B).

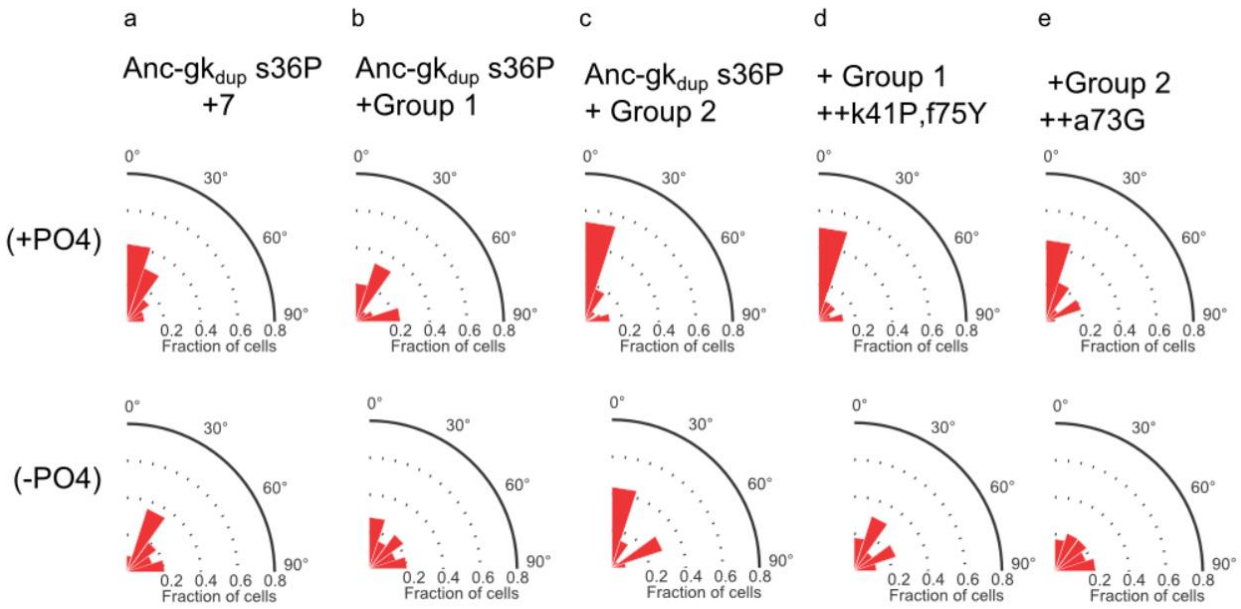


Figure 16. Screening of Anc-GK1$_{PID}$ substitutions in the Anc-gk$_{dup}$ enzyme for conferral of phospho-specific spindle orientation to the non-specific s36P. Seven candidate substitutions (**A**) conferred moderate phospho-specificity, but not to levels of extant controls (see Fig.6). A subsection of the seven candidates (group 1) ablated all spindle orientation (**B**) while an additional group (group 2) had no effect (**C**). Two substitutions of group 2 conferred phospho-specific spindle orientation when combined with group 1 substitutions (**D**). One substitution from group 1 was sufficient to confer phospho-specificity when combined with the mutations from group 2.

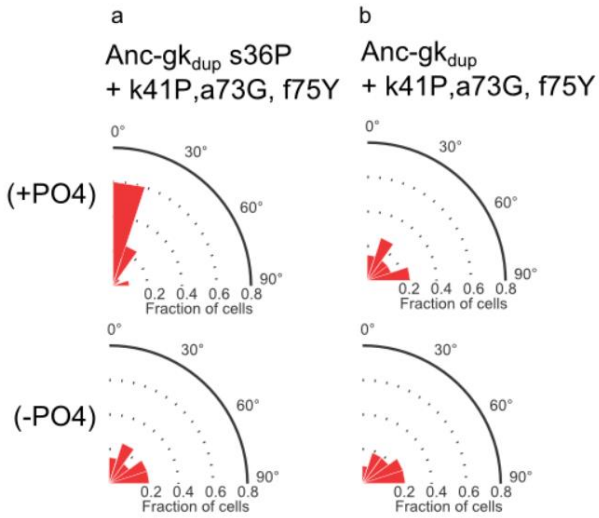


Figure 17. Anc-GK1$_{pid}$ minimal phospho-specific substitutions in Anc-gk$_{dup}$. Substitutions k41P, a73G and f75Y confer phospho-specific spindle orientation to Anc-gk$_{dup}$ s36P (**A**) but have no activity without s36P (**B**).

The above results suggested a potential mechanism whereby these residues enhance that of s36P. We noted that a73G in combination with s36P and group 1 substitutions ablated spindle orientation activity (Figure 16E) and hypothesized that this loss of spindle orientation capability may be due to a lowered Pins binding affinity. Furthermore, if the other substitutions k41P and f75Y possibly increased affinity for phospho-Pins only, the combined effect would be to exclude unphosphorylated binding while maintaining or improving phosphorylated binding. To test this hypothesis, we tested the effects of these substitutions with fluorescence anisotropy to better understand their effects on the GK / Pins interaction. Surprisingly, I found that the combined substitutions actually weakened overall phospho-Pins affinity by ~100uM, while weakening unphosphorylated Pins by ~300uM, suggesting these substitutions together preferentially exclude unphosphorylated interactions while leaving the phosphorylated interaction roughly intact (Table 1). Even more interesting was the observation that these affinity for phospho-Pins was very weak at over half a milimolar and yet was still able to orient the mitotic spindle while the unphosphorylated affinity was unable to do so. From this observation it may be possible that there is a thresholding effect within spindle orientation assays where affinities can be tolerated up to some unknown affinity, after which where they no longer function. With this in mind, the combined substitutions would appear to just barely push unphosphorylated Pins over this functional threshold, excluding the unphosphorylated interactions. Further attempts to dissect the molecular effects of these mutations were inconclusive; adding k41P and f75Y alone moderately improved affinities to both phosphorylated and unphosphorylated Pins while a73G alone had little effect. While more work remains to be done to elucidate the epistatic effects and mechanisms by which these substitutions work,

the functional data and anisotropy data suggest a mechanism by which the unphosphorylated interaction is preferentially excluded.

Construct (N=3)	phospho-Pins (μM)	Pins (μM)
Anc-gk_{dup}	N/D	N/D
Anc-gk_{dup} s36P	454.7 ± 40.7	439.0 ± 53.15
Anc-gk_{dup} s36P +k41P,a73G,f75Y	564.1 ± 34.33	748.6 ± 27.02
Anc-$GK1_{pid}$	27.94 ± 0.065	N/D
Drosophila Dlg-GK domain	0.9367 ± 0.1089	N/D

Table 1. Table of fluorescence anisotropy results for Anc-gk_{dup} constructs to phosphorylated and unphosphorylated Pins peptides as described previously. Anc-gk_{dup} alone does not bind either Pins peptide, but binds peptides with similar affinities upon addition of s36P. Further addition of phospho-specificity substitutions of k41P,a73G, f75Y preferentially weaken affinity for unphosphorylated Pins, but are not high affinity like Anc-$GK1_{PID}$ or the extant drosophila Dlg-GK domain. Each affinity is the mean of three runs with standard error. N/D – not detected.

While the above substitutions were sufficient to recapitulate phospho-specific spindle orientation in anc-gk_{dup}, the lack of strong binding affinity next led me to assay their necessity in the derived ancestor Anc-$GK1_{PID.}$ I used site-directed mutagenesis to revert these residues from their Anc-$GK1_{PID}$ identities and tested their effects by fluorescence anisotropy. As predicted, we found that reversion of the substitutions from Anc-$GK1_{PID}$ identities to Anc-gk_{dup} identities (P39k, G71a, Y73f) drastically reduced phospho-Pins affinity while increasing affinity for unphosphorylated-Pins (Table 2).

Reversion of the hinge proline (P34s) to the Anc-gk_{dup} serine identity also reduced phospho-Pins affinity and raised Pins affinity, and combining all four reversions resulted in the complete loss of phospho-specificity altogether with equal affinity for either phosphorylation state. Taken together, I conclude that these substitutions are necessary for phospho-specificity in the ancestral GK domain, and likely constitute essential residues in extant GK domain phospho-specificity.

Construct (N=3)	phospho-Pins (μM)	Pins (μM)
Anc-$GK1_{PID}$	27.94 ± 4.48	N/D
Anc-$GK1_{PID}$ P34s	343.4 ± 20.23	726.8 ± 68.92
Anc-$GK1_{PID}$ P39k, G71a, Y73f	231.2 ± 6.71	863.2 ± 40.94
Anc-$GK1_{PID}$ P34s P39k, G71a, Y73f	532.3 ± 51.37	527.6 ± 50.12

Table 2. Table of fluorescence anisotropy binding results for Anc-$GK1_{PID}$ reversion experiments. Reversion of the hinge proline to the Anc-gk_{dup} identity drastically reduces affinities for phosphorylated Pins, while increasing affinity for unphosphorylated Pins. Reversion of residues identified as phospho-specific in Anc-gk_{dup} identities (see Table 1) also decreases phosphorylated affinity while increasing unphosphorylated affinity. The combined loss of the hinge proline and phospho-specificity residues completely removes all phospho-specificity.

Although the exact functional mechanisms of the above phospho-specificity substitutions remain to be elucidated, it is clear that they constitute a significant portion of the GK domain's phospho-specificity evolution. Our findings that these substitutions may

function by destabilizing the unphosphorylated interaction while maintaining affinity for the phosphorylated interaction are in line with other biochemical systems, where removal of binding affinity of unwanted interactions is found instead of building of high affinity preferential interactions(45). Indeed, evolution may favor such mechanisms, as the addition of negative interactions is much easier than addition of highly-specific interactions.

Methods

Protein binding fluorescence anisotropy assay. Binding of *D. melanogaster* Pins was assayed by fluorescence anisotropy on a Tecan Sapphire plate reader equipped with automatic polarizers using default settings for anisotropy assays. Two FITC-labeled peptides (GVRVRRQ(pS)MEQLDLIKITPD, phosphorylated, and GVRVRRQSMEQLDLIKITPD, unphosphorylated, Genscript) of the fly Pins-Linker peptide (0.25 μM) were incubated with increasing concentrations of GK protein in phospho-buffered saline solution with 1mM DTT. A one-site binding model was used to fit the data and infer binding affinity in Graphpad Prism.

Spindle Orientation Assays. Maintenance of S2 cells, construction of expression plasmids, and cell-adhesion/spindle orientation assays were performed as detailed previously (17). S2 cells were transfected with constructs for FLAG- or HA-tagged GK protein and for Pins-GFP-Echinoid fusion protein, using Effectene reagent (Qiagen) with 0.4–1 μg total DNA for 24–48 h. Endogenous Dlg and Aurora A were knocked down using RNAi: transfected cells were incubated for 1 h in serum-free media containing approximately 1 μg RNAi followed by 72 h in normal growth media. Protein expression was then induced by adding 500 μM CuSO4 for 24 h. Cell-adhesion clustering was induced

by constant rotation at approximately 175 rpm for 1–3 h. Pins fusion protein was visualized by fluorescence (excitation 488 nm, emission 509 nm). Mitotic spindles were visualized using rat anti-tubulin (Abcam 1:500) and goat anti-rat conjugated to Alexa 555 (Life Technologies, 1:500). In addition, GK was visualized using mouse anti-FLAG or anti-HA (Sigma 1:500) and chicken anti-mouse:Alexa 647 (Life Technologies 1:200); histones were visualized using rabbit antiphospho-histone3 (Upstate 1:8000) and donkey anti-rabbit:Alexa 488 (Life Technologies, 1:500). For immunostaining, clustered cells were fixed in 4% paraformaldehyde for 20 min, washed, and incubated with primary antibodies overnight at 4 °C in buffer (0.01% saponin plus 0.1% albumin diluted in phosphate-buffered saline). Slides were subsequently washed and fluorescently linked secondary antibodies were added for 2 h at room temperature. Finally, slides were again washed and mounted using Vectashield Hardset medium (Vector Laboratories). All images were collected using a Biorad Radiance 2100 confocal microscope with a 1.4NA 60X objective . The spindle angles were measured for ~20 cells for each condition and displayed in a radial histogram.

CHAPTER IV

DISCUSSION

The findings from the previous chapters on the molecular evolution of the GK domain have broad implications beyond the molecular history of a protein interaction domain. GK domains and the MAGUK family as a whole are a widely functioning protein family involved in many crucial metazoan processes, including essential cellular processes such as spindle orientation, but also many higher functions such as neuronal synapse organization. The GK domain is an essential part of these processes and thus the elucidation of its evolutionary history and mechanisms contributes to the better understanding of metazoan processes. In this chapter I will discuss the implications from the previous two chapters in the light of the many MAGUK functions and their place in evolution, focusing on the evolution of spindle orientation pertaining to early metazoa, the establishment of phospho-specificity and more general GK domain evolution as pertaining to higher functions in other MAGUK GK domains.

Although it has been known for some time that the MAGUK GK domain is present throughout metazoa and more recently, choanoflagellates, an unexpected major result from chapter II is the finding that the original GK domain ancestor, Anc-GK1$_{PID}$, possessed spindle orientation activity. While Pins binding and spindle orientation are known functions of the Dlg family of GK domains, the other major MAGUK families possess very different cellular functions and are thus were considered unlikely to participate in Pins binding and spindle orientation. Because the Dlg family is only one of seven unique MAGUK families, it was unclear which molecular function would prove to be ancestral to the GK domain family. We observed that the Dlg-GK family functionality of Pins binding

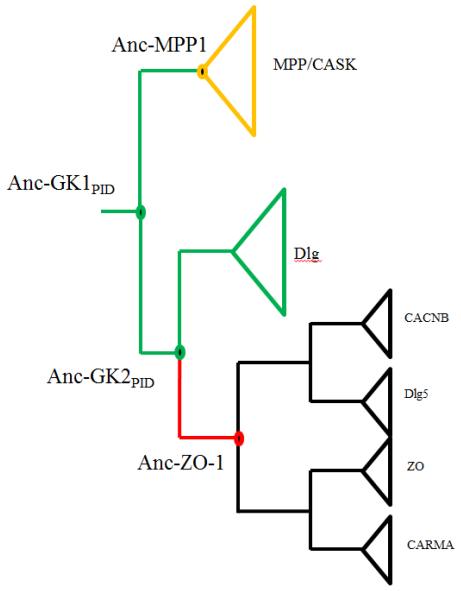
and spindle orientation was the ancestral function of the GK domain ancestor, suggesting that Pins-binding / spindle orientation capability was lost in other lineages or potentially maintained alongside other known functions.

This result indicates that Pins binding / spindle orientation is the original ancestral function, and that other functions would are derived from this original function. Furthermore, it is possible to observe this change in function by characterizing ancestors of these additional GK families and pin-pointing when spindle orientation / Pins binding was lost. Preliminary experiments along these lines of thought provide evidence that this may have been the case. Functional characterization of the ancestor to the MPP/CASK family indicates that this ancestor maintained Pins-binding ability similar to the progenitor Anc-$GK1_{PID}$ (Figure 18). Further derived ancestors likely lost Pins binding affinity in favor of known MPP family ligands as the family specialized. In contrast, characterization of the ancestor to the ZO/Dlg5/CARMA/CaCNB famly of GK domains found that all Pins-binding ability was lost completely. Interestingly, this loss of Pins-binding is accompanied by a 20 amino-acid deletion event in the Pins-binding region of the GK domain which is conserved in all derived ancestors and extant proteins. Analysis of extant ZO-like GK domains through crystallographic studies further shows that this initial deletion event has resulted in the complete loss of the GBD subdomain which supports Pins binding, indicating that these family members have completely lost all Pins-binding capability.

This preliminary result suggests that the ZO/Dlg5/CARMA/CACNB family function was created by a further neofunctionalizing event at the cost of Pins-binding and was further diversified into the four separate families. These results have potential importance for future studies of MAGUK GK domains. First, the finding that the

MPP/CASK family functionality is evolved from the previous Pins-binding / spindle orientation function suggests that residues important in Pins binding may also be relevant in MPP/CASK specific ligand binding. In the opposite way, the deletion of crucial Pins binding residues in the ZO/Dlg5 and related families suggests these domains bind protein ligands in a completely different manner, as they lack the GBD subdomain completely(50). Further studies will be necessary to determine the exact functional evolution of non Dlg-family GK domain molecular functions.

a



b

Construct (N=3)	phospho-Pins (μM)
Anc-GK1$_{PID}$	27.94 ± 4.48
Anc-GK2$_{PID}$	1.430 ± 0.08
Anc-MPP1	83.39 ± 6.71
Anc-ZO-1	N/D

Figure 18. Preliminary investigation of Pins-binding by nearby non-Dlg lineage ancestors. **A**. Summary schematic of Pins-binding ability(green) in GK domain phylogeny. Pins binding is first present in the GK domain ancestor Anc-GK1$_{PID}$ and is maintained in the Dlg family lineage and is also present in the ancestor to all MPP family(yellow) GK domains, Anc-MPP1. Pins binding is lost(red) in the transition from Anc-GK2$_{PID}$ to the ancestor of all ZO, Dlg5, CARMA and CACNB family GK domains, Anc-ZO-1. **B**. Summary of phospho-Pins disassociation constants(K_d,μM) by fluorescence anisotropy for the ancestors in panel A.

In addition to implications regarding non-Dlg families, the finding that Pins-binding / spindle orientation evolved as the first GK domain functionality also has major

implications for the evolution of mitotic spindle orientation. From the GK domain phylogeny, Anc-GK1$_{PID}$, the first GK domain ancestor, diverged from the guanylate kinase ancestor Anc-gk$_{dup}$ before the divergence of choanoflagellates and metazoans. Although it has been known that spindle orientation is necessary for organized tissues and cellular differentiation, the finding that spindle orientation capability evolved before metazoa further highlights spindle orientation itself as even more ancient than previously thought. Further evidence from chapter II that modern choanoflagellates also orient their spindle to generate spatially organized colonies and may also have functioning Pins/GK/KHC-73 homologs suggests that Pins/GK pathway itself may have evolved before metazoa, suggesting the importance of these molecules in the building of true animal multicellularity.

Lastly, the causal historical substitutions themselves which neofunctionalized the ancient guanylate kinase ancestor to that of a phospho-peptide binding and spindle orientation domain have interesting implications for molecular mechanisms in evolution. Unlike many instances of neofunctionalization in protein evolution where variations of the same functionality are evolved via stepwise modifications of binding sites or active sites, the GK domain may have undergone a relatively rapid and sequentially simple evolution that did not a stepwise evolution of a novel binding site from scratch. The finding that a single amino acid substitution is responsible for a substantial change in function suggests that evolution has made use of existing binding sites and repurposed them for protein binding by modifying protein biophysics, such as conformational dynamics, instead of a step-by-step evolution of a new binding site. Further substitutions as discussed in chapter III likely refined the initial latent binding site exposed by s36P and functioned to exclude

unphosphorylated binding. Future studies are necessary to resolve the exact biochemical mechanism on how the described substitutions exactly function, however the current work described here establishes at the minimum the genetic changes responsible for the evolution of the GK protein interaction domain.

Although future studies are necessary to elucidate the exact biophysical mechanisms of the GK domain's historical hinge and phospho-specificity substitutions, previous extant studies have shown modifying protein dynamics and conformational bias could be sufficient to alter GK enzyme functionality to give rise to GK domain-like function (22). Our finding that these same alterations took place in the actual evolutionary history of the evolution of the GK domain further supports the growing idea that modifications to protein conformational dynamics themselves can be an effective evolutionary mechanism in the evolution of novel protein functions.

APPENDIX

ANCESTRAL RECONSTRUCTION TABLES

Residue No.	Reconstructed Amino Acid	PP	Alt State	PP
1	A	0.92	K	0.03
2	P	0.87	S	0.08
3	R	1.00		
4	P	1.00		
5	V	0.95	I	0.04
6	V	1.00		
7	L	0.96	I	0.04
8	S	0.99	C	0.01
9	G	1.00		
10	P	1.00		
11	S	1.00		
12	G	1.00		
13	S	0.98	T	0.02
14	G	1.00		
15	K	1.00		
16	S	1.00		
17	T	1.00		
18	L	1.00		
19	L	0.99	I	0.01
20	K	1.00		
21	K	0.50	R	0.50
22	L	1.00		
23	L	0.51	F	0.46
24	K	0.92	Q	0.03
25	E	1.00		
26	F	0.98	Y	0.02
27	P	1.00		
28	D	0.99	N	0.01
29	E	0.51	K	0.17
30	F	1.00		
31	G	1.00		
32	F	1.00		
33	S	1.00		
34	V	1.00		
35	S	1.00		
36	H	1.00		
37	T	1.00		
38	T	1.00		
39	R	1.00		
40	K	1.00	Q	0.00
41	P	1.00		
42	R	1.00		
43	P	1.00		
44	G	1.00		
45	E	1.00		
46	V	1.00	I	0.00
47	N	0.99	D	0.01
48	G	1.00		
49	K	0.88	R	0.12
50	D	1.00	E	0.00
51	Y	1.00		
52	Y	0.63	H	0.37
53	F	1.00		
54	V	0.99	T	0.00
55	T	1.00	S	0.01
56	R	1.00	K	0.00
57	E	0.93	D	0.07
58	E	1.00	D	0.00
59	M	1.00		
60	E	1.00		
61	Q	0.75	K	0.14
62	A	0.78	M	0.20
63	I	1.00		
64	E	0.86	K	0.09
65	K	0.97	E	0.02
66	G	1.00		
67	E	1.00	D	0.00
68	F	1.00		
69	I	1.00		
70	E	1.00		
71	H	0.85	Y	0.10
72	A	1.00		
73	E	0.99	Q	0.01
74	F	0.99	Y	0.02
75	S	1.00		
76	G	1.00		
77	N	1.00		
78	L	0.99	M	0.01
79	Y	1.00		
80	G	1.00		
81	T	1.00		
82	S	1.00		
83	K	1.00		

Residue No.	Reconstructed Amino Acid	PP	Alt State	PP
84	K	0.99	Q	0.01
85	A	1.00		
86	V	1.00		
87	Q	0.84	K	0.08
88	D	0.99	E	0.01
89	V	1.00		
90	Q	0.89	L	0.03
91	S	0.76	A	0.10
92	Q	0.91	K	0.07
93	G	1.00		
94	K	1.00	R	0.00
95	I	1.00	V	0.00
96	C	1.00		
97	I	1.00	V	0.00
98	L	1.00		
99	D	1.00		
100	I	1.00		
101	D	1.00		
102	M	1.00		
103	Q	1.00		
104	G	1.00		
105	V	1.00		
106	K	1.00	R	0.00
107	N	0.66	Q	0.22
108	I	0.99	V	0.01
109	K	1.00		
110	K	1.00	E	0.00
111	T	1.00		
112	D	1.00	E	0.00
113	L	1.00		
114	N	1.00		
115	P	1.00		
116	I	0.54	L	0.45
117	Y	1.00	F	0.00
118	I	0.98	V	0.02
119	F	1.00		
120	I	1.00		
121	Q	0.80	K	0.18
122	P	1.00		
123	P	1.00		
124	S	1.00		
125	M	0.62	I	0.23
126	E	1.00		
127	E	1.00	A	0.00
128	L	1.00		
129	E	1.00		
130	K	1.00		
131	R	1.00		
132	L	1.00		
133	R	1.00		
134	G	0.88	S	0.11
135	R	1.00		
136	G	1.00	N	0.00
137	T	1.00		
138	E	1.00	D	0.00
139	T	1.00		
140	E	1.00		
141	E	1.00		
142	S	1.00		
143	L	1.00		
144	Q	0.79	R	0.12
145	K	1.00		
146	R	1.00		
147	L	1.00		
148	A	0.93	S	0.04
149	T	0.60	A	0.39
150	A	1.00		
151	K	0.64	R	0.35
152	E	0.62	K	0.13
153	E	1.00		
154	M	0.94	L	0.06
155	E	1.00		
156	Y	1.00		
157	G	0.70	A	0.26
158	K	0.60	Q	0.34
159	E	0.48	T	0.27
160	P	0.99	E	0.00
161	G	1.00		
162	A	0.49	S	0.35
163	F	1.00		
164	D	1.00		
165	H	0.95	K	0.03
166	I	0.63	V	0.37

Residue No.	Reconstructed Amino Acid	PP	Alt State	PP
167	I	0.99	V	0.01
168	V	1.00	I	0.00
169	N	1.00		
170	D	1.00		
171	D	0.99	N	0.01
172	L	1.00		
173	E	0.79	D	0.21
174	K	0.82	E	0.11
175	A	1.00		
176	Y	1.00		
177	E	0.91	Q	0.03
178	E	0.69	K	0.29
179	L	1.00		
180	K	1.00	R	0.00
181	D	0.64	E	0.29
182	F	0.98	I	0.02
183	I	0.98	L	0.01
184	I	0.48	M	0.30
185	Q	0.55	E	0.40
186	E	1.00	Q	0.00
187	K	0.69	I	0.20

Mean Posterior Probability	Number Expected Errors
0.94	11.72

Figure A1. The reconstructed sequence of Anc-gk_{dup} is shown with the maximum likelihood and next-best amino acid state shown at each site, along with their posterior probabilities (PP). Ambiguously reconstructed sites with a next-best state with PP > 0.2 are shaded gray. The mean posterior probability over sites and the number of expected errors in the complete ML sequence given the posterior probabilities at each site are shown.

Residue No.	Reconstructed Amino Acid	PP	Alt State	PP
1	R	0.21	A	0.17
2	P	0.36	K	0.26
3	R	1.00	K	0.00
4	P	1.00		
5	V	0.75	I	0.23
6	V	0.91	I	0.09
7	L	0.91	I	0.08
8	L	0.48	I	0.41
9	G	1.00		
10	P	1.00		
11	S	0.82	P	0.13
12	V	0.65	T	0.23
13	G	1.00		
14	K	0.92	R	0.08
15	N	0.66	D	0.22
16	E	0.62	Q	0.10
17	L	0.96	I	0.04
18	K	0.42	N	0.17
19	K	0.35	N	0.20
20	R	0.67	K	0.33
21	L	1.00		
22	I	0.85	L	0.08
23	A	0.22	E	0.22
24	E	0.85	D	0.05
25	F	0.72	Y	0.24
26	P	1.00		
27	D	0.84	E	0.14
28	R	0.42	K	0.29
29	F	1.00	Y	0.00
30	G	1.00	A	0.00
31	S	0.88	T	0.07
32	A	0.39	S	0.38
33	V	0.99	I	0.01
34	P	0.99	S	0.01
35	H	1.00		
36	T	1.00		
37	T	1.00	S	0.00
38	R	1.00		
39	P	0.97	S	0.01
40	P	1.00	R	0.00
41	R	1.00	K	0.00
42	P	0.84	E	0.05
43	G	1.00		
44	E	1.00		
45	V	0.92	I	0.07
46	D	0.73	N	0.23
47	G	1.00		
48	R	0.79	K	0.19
49	D	0.99	E	0.01
50	Y	1.00		
51	H	0.89	Y	0.11
52	F	1.00		
53	V	1.00	I	0.00
54	S	0.80	T	0.20
55	R	0.99	K	0.01
56	E	0.94	Q	0.04
57	Q	0.40	E	0.38
58	M	1.00		
59	E	1.00	D	0.00
60	A	0.25	K	0.25
61	D	0.89	G	0.04
62	I	1.00	V	0.00
63	E	0.34	Q	0.31
64	N	0.26	D	0.21
65	G	0.99	N	0.01
66	K	0.49	E	0.39
67	F	1.00	Y	0.00
68	I	0.97	V	0.02
69	E	1.00		
70	H	0.53	Y	0.32
71	G	0.99	A	0.01
72	E	0.87	Q	0.15
73	Y	0.98	F	0.02
74	K	0.47	E	0.15
75	G	0.97	D	0.01
76	N	1.00	H	0.00
77	L	1.00	I	0.00
78	Y	1.00		
79	G	1.00		
80	T	1.00		
81	S	1.00		
82	I	0.78	V	0.12
83	D	0.38	E	0.27
84	A	0.66	S	0.34
85	V	1.00	I	0.00
86	R	0.83	Q	0.13
87	D	0.70	N	0.10
88	V	1.00	I	0.00
89	I	0.47	V	0.47
90	N	0.20	E	0.19
91	Q	0.72	K	0.10
92	G	1.00		
93	K	0.99	R	0.01
94	I	0.58	V	0.39
95	C	1.00		
96	I	0.90	V	0.10
97	L	1.00		
98	D	1.00		
99	V	0.84	I	0.13
100	N	0.38	S	0.25
101	A	0.45	P	0.43
102	Q	0.92	H	0.06
103	A	0.99	G	0.01
104	I	0.52	V	0.32
105	K	0.94	E	0.05
106	R	0.19	K	0.16
107	L	0.94	I	0.02
108	R	0.67	K	0.27
109	T	0.37	K	0.21
110	A	0.72	S	0.14
111	E	0.57	D	0.33
112	L	1.00	F	0.00
113	N	0.46	K	0.22
114	P	1.00		
115	I	0.78	L	0.10
116	V	0.93	I	0.06
117	I	0.93	V	0.07
118	F	1.00		
119	I	0.99	V	0.01
120	K	0.80	Q	0.10
121	P	1.00		
122	P	0.99	A	0.01
123	S	1.00		
124	L	0.47	I	0.23
125	E	0.91	D	0.07
126	S	0.29	K	0.20
127	L	0.96	I	0.04
128	M	0.39	L	0.22
129	K	0.38	R	0.30
130	S	0.16	T	0.13
131	N	0.66	S	0.17
132	K	0.30	P	0.25
133	Q	0.48	R	0.11
134	L	0.43	M	0.30
135	T	0.97	S	0.03
136	E	0.93	D	0.07
137	E	1.00	D	0.00
138	S	0.41	T	0.35
139	L	0.95	A	0.02
140	R	0.81	Q	0.11
141	K	0.88	E	0.10
142	V	0.38	M	0.34
143	L	0.37	F	0.27
144	D	0.52	E	0.41
145	T	0.46	S	0.15
146	A	1.00	S	0.00
147	Q	0.59	K	0.14
148	K	0.66	E	0.20
149	M	0.90	L	0.09
150	E	1.00	Q	0.00
151	Q	0.79	K	0.07
152	E	0.70	D	0.15
153	Y	0.98	F	0.02
154	G	0.79	A	0.10
155	H	0.97	Q	0.03
156	Y	0.62	F	0.25
157	F	1.00		
158	D	0.98	E	0.01
159	A	0.25	H	0.24
160	V	0.86	I	0.14
161	I	0.85	V	0.15
162	V	0.92	I	0.05
163	N	1.00		
164	D	0.99	E	0.01
165	D	0.95	N	0.04
166	L	1.00	M	0.00
167	D	0.82	E	0.17
168	E	0.49	K	0.25
169	A	0.95	T	0.05
170	Y	0.99	F	0.01
171	E	0.30	N	0.18
172	E	0.31	Q	0.29
173	L	0.92	V	0.06
174	K	0.72	R	0.12
175	E	0.55	D	0.20
176	A	0.42	V	0.26
177	I	0.93	V	0.06
178	N	0.22	R	0.17
179	K	0.41	R	0.26
180	Q	0.78	E	0.11
181	S	0.31	K	0.26

Mean Posterior Probability	Number Expected Errors
0.77	41.32

Figure A2. The reconstructed sequence of Anc-GK1$_{PID}$ is shown with the maximum likelihood and next-best amino acid state shown at each site, along with their posterior probabilities (PP). Ambiguously reconstructed sites with a next-best state with PP > 0.2 are shaded gray. The mean posterior probability over sites and the number of expected errors in the complete ML sequence given the posterior probabilities at each site are shown.

Species Name	Common Name	Phylum	GK Family	Abbreviation	Accession #	Database
Schistosoma mansoni	blood fluke	Invertebrata, Platyhelminthes	CACNB	SchistCACNB	CCD77788.1	Genbank
			ZO	SMansoniZO	CD129964.1	Genbank
Trichoplax adhaerens	Trichoplax	Parazoa, Placozoa	CACNB	TrichoCACNCB	XP_002110341.1	Genbank
			GMP Kinase	TrichoGUK	XP_002110408.1	Genbank
			MPP	TrichoMPP	XP_002110426.1	Genbank
			Dlg	TrichoDLG	XP_002108800.1	Genbank
			Dlg5	TrichoDLG5	XP_002112079.1	Genbank
Cyanea capillata	Lion's mane jellyfish	Cnidaria	CACNB	CyanCACNB1	AAB87751.1	Genbank
Hydra vulgaris	fresh water polyp	Cnidaria	CACNB	HydraCACNB	XP_002156135.2	Genbank
			ZO	HydraZO1	AAK38322.1	Genbank
Physalia physalis	portuguese man o' war	Cnidaria	CACNB	PhysCACNB1	ABD59026.1	Genbank
Nematostella vectensis	starlet sea anemone	Cnidaria	CACNB	NemoCACNB1	XP_001626057.1	Genbank
			CASK	NemosCASK	XP_001631427.1	Genbank
			Dlg	Nematostel	XP_001638123.1	Genbank
Ciona intestinalis	vase tunicate	Chordata, tunicata	CACNB	CionaCACNB	XP_002129248.1	Genbank
			GMP Kinase	CionaGUK	XP_002120364.1	Genbank
			MPP	CionaMPP5	XP_002123052.2	Genbank
			CASK	CionaCASK	XM_004225487.1	Genbank
			MPP	CionaMPP7	XP_002122568.2	Genbank
			Dlg	CionaDlg1	XP_002131501.1	Genbank
			ZO	CionaZO	XP_002129483.1	Genbank
Schistosoma japonicum	blood fluke	Invertebrata, Platyhelminthes	CACNB	SchistCACNB1	AAX28385.2	Genbank
Strongylocentrotus purpuratus	purple sea urchin	Echinodermata	CACNB	StroCACNB1	XP_790453.3	Genbank
			MPP	StrongyMPP	XP_784409.3	Genbank
			CASK	StronCASK	XP_003726866.1	Genbank
			Dlg	Strongyloc	XP_797198.3	Genbank
			Dlg5	UrchinDLG4	XP_003730681.1	Genbank
			ZO	StrongyZO1	XP_782687.3	Genbank
Caenorhabditis elegans	nematode	Nematoda	CACNB	WormCACNB1	NP_491193.2	Genbank
			MPP	CelegansMP	NP_505265.1	Genbank
			CASK	WormCASK	NP_001024588.1	Genbank
			ZO	CelegansZO	NP_001252378.1	Genbank
Heterololigo bleekeri	Inshore squid	Mollusca	CACNB	LoliCACNB1	BAB88219.1	Genbank
Lymnaea stagnalis	great pond snail	Mollusca	CACNB	LymnCACNB1	AEE65419.1	Genbank
			CASK	LymnCASK	AAO83853.1	Genbank
Mus musculus	house mouse	Vertebrata, mammalia	CACNB	MusCACNB1	EDL16111.1	Genbank
			CARMA	MusCARMA3	P58660.1	Genbank
			CARMA	MusCARMA1	BAC36069.1	Genbank
			CARMA	MusCARMA2	XP_006532440.1	Genbank
			GMP Kinase	MusGUK	NP_001152882.1	Genbank
			MPP	MusMPP4	AAH61694.2	Genbank
			MPP	MusMPP7	XP_006526361.1	Genbank
			MPP	MusMPP3	XP_006532303.1	Genbank
			MPP	MusMPP5	XP_006516162.1	Genbank
			MPP	MusMPP6	AAD45009.1	Genbank
			MPP	MusMPP2	XP_006533750.1	Genbank
			CASK	MusCASK	CAA76647.1	Genbank
			MPP	MusMPP1	NP_032647.1	Genbank
			Dlg	MouseDlg3	NP_001171251.1	Genbank
			Dlg	MouseDlg4	EDL12524.1	Genbank
			Dlg	MouseDlg1	AAC31653.1	Genbank
			Dlg	MouseDlg2	NP_001229976.1	Genbank
			Dlg5	Mus	BAB30534.1	Genbank
			ZO	MusZO1	BAA03274.1	Genbank
			ZO	MusZO2	NP_035727.2	Genbank
			ZO	MusZO3	XP_006513772.1	Genbank
Drosophila melanogaster	fruit fly	Arthropoda	CACNB	FlyCACNB1	NP_001260377.1	Genbank
			GMP Kinase	FlyGUK	NP_648408.1	Genbank
			MPP	FlyMPP5	AAR82754.1	Genbank
			CASK	FlyCASK	NP_732662.2	Genbank
			MPP	FlyMPP7	NP_001286309.1	Genbank
			Dlg	FlyDlg1	AAL39553.1	Genbank
			Dlg5	flyDLG5	NP_001260385.1	Genbank
			Dlg5	FlyZO1	AAK93389.1	Genbank
Tribolium castaneum	red flour beetle	Arthropoda	CACNB	TriboCACNB	XP_008194397.1	Genbank
Apis mellifera	western honey bee	Arthropoda	CACNB	BeeCACNB1	XP_006560409.1	Genbank
			MPP	ApisMPP	XP_006570377.1	Genbank
			Dlg5	ApisMelDLG	XP_006569218.1	Genbank
Homo sapiens	human	Vertebrata, mammalia	CARMA	HomoCARMA3	BAH12273.1	Genbank
			CARMA	HomoCARMA1	AAL34460.1	Genbank
			CARMA	HomoCARMA2	NP_077015.2	Genbank
			GMP Kinase	HomoGUK	AAQ02435.1	Genbank
			MPP	HomoMPP4	EAW70289.1	Genbank
			MPP	HomoMPP7	EAW86046.1	Genbank
			MPP	HomoMPP3	XP_006721980.1	Genbank
			MPP	HomoMPP5	CAH56281.1	Genbank
			MPP	HomoMPP6	XP_005249832.1	Genbank
			MPP	HomoMPP2	BAG61470.1	Genbank
			CASK	HomoCASK	1KGD_A	Genbank
			MPP	HomoMPP1	3NEY_A	Genbank
			Dlg	homoDlg3	NP_001159750.1	Genbank
			Dlg	HomoDlg4	AAH40533.1	Genbank
			Dlg	HomoDlg1	CAD38582.2	Genbank
			Dlg	HomoDlg2	XP_003129756.3	Genbank
			Dlg5	HomoDLG5	BAG37308.1	Genbank
			ZO	HomoZO2	AAD56219.2	Genbank
			ZO	HomoZO3	BAB55020.1	Genbank
Tetraodon nigroviridis	green spotted puffer	Vertebrata	CARMA	TetraCARMA	CAF95531.1	Genbank
			CARMA	TetraoCARM	CAF91731.1	Genbank
			MPP	TetraoMPP4	CAG10306.1	Genbank
			MPP	TetraoMPP7	CAF98752.1	Genbank
			MPP	TetraoMPP5	CAG03545.1	Genbank
			MPP	TetraoMPP2	CAG03075.1	Genbank
			CASK	TetraCASK	CAG06710.1	Genbank
			MPP	TetraoMPP1	CAG01565.1	Genbank
			Dlg	TetraoDLG3	CAG09800.1	Genbank
			Dlg	TetraDlg4	CAG00581.1	Genbank
			Dlg5	tetraodonD	CAF98636.1	Genbank
			ZO	TetraoZO3	CAG04855.1	Genbank
Danio rerio	zebrafish	Vertebrata	CARMA	DanioCARMA	XP_685144.6	Genbank
			GMP Kinase	DanioGUK	XP_005166853.1	Genbank
			MPP	DanioMPP4	XP_684526.5	Genbank
			MPP	DanioMPP7	NP_571051.1	Genbank
			MPP	DanioMPP3	XP_700838.5	Genbank
			MPP	DanioMPP5	Q8JHF4.2	Genbank
			MPP	DanioMPP2	NP_001103181.1	Genbank
			MPP	DanioMPP6	AAI71539.1	Genbank
			CASK	DanioCASK	NP_694420.1	Genbank
			MPP	DanioMPP1	NP_998857.1	Genbank
			Dlg	DanioDlg3	XP_005173330.1	Genbank
			Dlg	DanioDlg4	NP_998893.1	Genbank
			Dlg	DanioDlg1	XP_005165873.1	Genbank
			Dlg	DanioDlg2	XP_005172761.1	Genbank
			Dlg5	DanioDLG5	XP_005156738.1	Genbank
			ZO	DanioZO1	XP_005168993.1	Genbank
			ZO	DanioZO2	NP_001188500.1	Genbank
			ZO	DanioZO3	XP_005170498.1	Genbank
Xenopus laevis	african clawed frog	Vertebrata	CARMA	XenoCARMA1	XM_002937309.2	Genbank
			CARMA	XenoCARMA2	XP_002937355.2	Genbank
			GMP Kinase	XenoGUK	AAH64272.1	Genbank
			MPP	XenoMPP7	NP_001106472.1	Genbank
			MPP	XenoMPP5	XP_002937246.1	Genbank
			MPP	XenoMPP6	NP_001017060.1	Genbank
			MPP	XenoMPP2	NP_001096176.1	Genbank
			CASK	XenoCASK	BC171041.1	Genbank
			MPP1	XenoMPP1	NP_001008152.1	Genbank
			Dlg	XenoDlg3	XP_004916885.1	Genbank
			Dlg	XenoDlg4	AAI35455.1	Genbank
			Dlg	XenoDlg1	NP_001039116.1	Genbank
			Dlg	XenoDlg2	XP_002942127.2	Genbank
			Dlg5	XenoDLG5	XP_004916077.1	Genbank
			ZO	XenoZO1	XP_004917990.1	Genbank
			ZO	XenoZO2	XP_004910834.1	Genbank
			ZO	XenoZO3	XP_002941439.2	Genbank
Gallus gallus	chicken	Vertebrata	CARMA	GallusCARM	NP_001006161.2	Genbank
			MPP	GallusMPP4	XP_001233047.2	Genbank
			MPP	GallusMPP7	XP_418583.2	Genbank
			MPP	GallusMPP5	NP_001186634.1	Genbank
			MPP	GallusMPP6	XP_418721.3	Genbank
			MPP	GallusMPP2	XP_004948695.1	Genbank
			CASK	GallusCASK	XP_005012538.1	Genbank
			MPP	GallusMPP1	NP_001007918.1	Genbank
			Dlg	GallusDlg3	XP_426264.4	Genbank
			Dlg	GallusDlg1	XP_004943379.1	Genbank
			Dlg	GallusDlg2	XP_004938975.1	Genbank
			Dlg5	GallusDLG5	XP_421604.3	Genbank
			ZO	GallusZO2	NP_990249.1	Genbank
			ZO	GallusZO3	XP_003642888.2	Genbank
Monodelphis domestica	short-tailed opossum	Vertebrata, mammalia	CARMA	PossomCARM	XP_007488475.1	Genbank
			Dlg	Possum2	XP_007490942.1	Genbank
Helicobacter pylori		Bacteria	GMP Kinase	HelicoGUK	WP_001058031.1	Genbank
Candidatus Atelocyanobacterium thalassa	cyanobacterium	Cyanobacteria	GMP Kinase	CyanobGUK	WP_012954113.1	Genbank
Synechococcus elongatus		Cyanobacteria	GMP Kinase	SyneGUK	WP_011242672.1	Genbank
Brachyspira hyodysenteriae		Spirochaetes	GMP Kinase	BrachyGUK	WP_012670217.1	Genbank
Lepeophtheirus salmonis	salmon louse	Arthropoda	GMP Kinase	LepeoGUK	ACO12190.1	Genbank
Branchiostoma floridae	florida lancelet	Chordata, cephalochordata	GMP Kinase	BranchGUK	XP_002605108.1	Genbank
			MPP	BranchMPP5	XP_002610256.1	Genbank
			MPP	BranchMPP2	XP_002588010.1	Genbank
			Dlg	BranchioDL	XP_002588761.1	Genbank
Neurospora crassa	red bread mold	Ascomycota	GMP Kinase	NeuroGUK	EAA33326.2	Genbank
Amphimedon queenslandica		Porifera	GMP Kinase	AmphiGUK	XP_003387740.1	Genbank
			Dlg	AmphiDLG	XP_003391362.1	Genbank
Monosiga brevicollis		Filozoa, choanoflagelleta	GMP Kinase	MonoGUK	XP_001748709.1	Genbank
			Dlg	MonosigaDL	XP_001743865.1	Genbank
Bombyx mori	silkworm	Arthropoda	GMP Kinase	BomGUK	NP_001040302.1	Genbank
Homarus americanus	american lobster	Arthropoda	GMP Kinase	LobsterGUK	EX471111.1	Genbank
Talaromyces marneffei ATCC 18224	penicillium	Ascomycota	GMP Kinase	PeniGUK	XP_002151796.1	Genbank
Candida albicans WO-1		Ascomycota	GMP Kinase	CandiGUK	EEQ46401.1	Genbank
Zygosaccharomyces rouxii CBS 732	penicillium	Ascomycota	GMP Kinase	Zygosaccha	XP_002497027.1	Genbank
Komagataella pastoris	pichia pastoris	Ascomycota	GMP Kinase	PichiaGUK	XP_002491978.1	Genbank
Aspergillus fumigatus Af293		Ascomycota	GMP Kinase	Apsergillu	XP_752351.1	Genbank
Kluyveromyces lactis NRRL Y-1140		Ascomycota	GMP Kinase	Kluyveromy	XP_451606.2	Genbank
Saccharomyces cerevisiae S288c	brewer's yeast	Ascomycota	GMP Kinase	YeastGUK	NP_010742.1	Genbank
Aplysia californica	sea slug	Mollusca	GMP Kinase	AplysiaGUK	EB325482.1	Genbank
			Dlg	Aplysiacal	NP_001191484.1	Genbank
Saccoglossus kowalevskii	acorn worm	Hemichordata	GMP Kinase	SaccoGUK	XP_002737679.1	Genbank
			Dlg	Saccogloss	ADB22636.1	Genbank
Capitella teleta	polychaete worm	Annelida	GMP Kinase	CapitellaG	ELU10045.1	Genbank
			Dlg	CapitellaC	ELU16460.1	Genbank
			Dlg5	CaptellaD	ELT92714.1	Genbank
			ZO	CapitellaZ	ELU09381.1	Genbank
Eptatretus stoutii	pacific hagfish		GMP Kinase	HagfishGUK	BJ650728.1	Genbank
Lethenteron camtschaticum	arctic lamprey		GMP Kinase	LampreyGUK	DC615446.1	Genbank
Squalus acanthias	dogfish shark	Chordata, Chondrichthyes	GMP Kinase	DogfishGUK	ctg22149	Skatebase
			Dlg5	DogfishDLG	ctg15832	Skatebase
Agrobacterium		Proteobacteria	GMP Kinase	AgroGUK	WP_007702513.1	Genbank
Sinorhizobium	rhizobia	Proteobacteria	GMP Kinase	SinoGUK	WP_011974957.1	Genbank
Halothermothrix orenii		Firmicutes	GMP Kinase	HaloGUK	WP_012635923.1	Genbank
Escherichia coli		Proteobacteria	GMP Kinase	EcoliGUK	WP_001618632.1	Genbank
Shigella boydii		Proteobacteria	GMP Kinase	ShigellaGU	WP_004996980.1	Genbank
Enterobacter sp. 638		Proteobacteria	GMP Kinase	Enterobact	WP_011915359.1	Genbank
Salmonella enterica		Proteobacteria	GMP Kinase	Salmonella	WP_000046969.1	Genbank
Neisseria subflava		Proteobacteria	GMP Kinase	NeisseriaG	WP_004519394.1	Genbank
Desulfurobacterium thermolithotrophum		Aquificae	GMP Kinase	DesulfGUK	WP_013637795.1	Genbank
Thermovibrio ammonificans		Aquificae	GMP Kinase	Thermovibr	WP_013538151.1	Genbank
Loa Loa	eye worm	Nematoda	CASK	LoaLoaCASK	XP_003143929.1	Genbank
Leucoraja erinacea	little skate	Chordata, Chondrichthyes	CASK	LeucCASK	ctg67297	Skatebase
			Dlg	LeucoDlg1	ctg12488	Skatebase
			Dlg	Skate2	ctg22079	Skatebase
Dasypus novemcinctus	nine-banded armidillo	Vertebrata, mammalia	MPP	ArmidMPP	XP_004480905.1	Genbank
Capsaspora owczarzaki		Opisthokonta, Filasterea	Dlg	Capsaspora	XP_004349749.1	Genbank
Salpingoeca rosetta		Opisthokonta, choanoflagelleta	Dlg	Srosetta	XP_004997111.1	Genbank
Oscarella lobularis	flesh sponge	Porifera	Dlg	OscarellaD	AM763386.1	Genbank
Sus scrofa	wild boar	Chordata, mammalia	Dlg	PigDLG3	XP_003484170.1	Genbank
			Dlg	PigDLG1	XP_005897895.1	Genbank
			Dlg	PigDLG2	XP_003129759.3	Genbank
Equus caballus	horse	Chordata, mammalia	Dlg	HorseDlg4	XP_001498988	Genbank
			Dlg	HorseDlg1	CAD38582.2	Genbank
			Dlg	HorseDlg2	XP_003365415.1	Genbank
Pediculus humanus corporis	body louse	Arthropoda	Dlg	Pediculus	XP_002432243.1	Genbank
Daphnia pulex	water flea	Arthropoda	Dlg	Daphniadlg	EFX75703.1	Genbank
			Dlg5	DaphniaDLG	EFX81288.1	Genbank
Idiosepius paradoxus	northern pigmy squid	Mollusca	Dlg	SquidDLG	DB915272.1	Genbank
Lottia gigantea	giant owl limpet	Mollusca	Dlg	Lottia	XP_009047619.1	Genbank
Alvinella pompejana	pompeii worm	Annelida	Dlg5	AlvinellaP	TERA13701	Alvinella pompejana project
Taeniopygia guttata	zebra finch	Chordata	Dlg5	FinchDLG5	XP_002192279.2	Genbank
Oncorhynchus mykiss	rainbow trout	Chordata	Dlg5	TroutDLG5	CDQ63988.1	Genbank
Camponotus floridanus	florida carpenter ant	Arthropoda	ZO	FlorAntZO	EFN63382.1	Genbank
Aedes aegypti	yellow fever mosquito	Arthropoda	ZO	EgyMosZO	XP_001656413.1	Genbank
Halocynthia roretzi	sea pineapple	Chordata, Tunicata	ZO	Halocynthi	AV386449.1	Genbank
Astyanax mexicanus	mexican tetra	Chordata	ZO	MinnowZO	XP_007246682.1	Genbank

Figure A3. Protein sequences with accessions and species names used in the phylogenetic analysis to generate Figure 5.

Species	Accession #
Homo sapiens	NP_037428.3
Mus musculus	XP_006502319.1
Ciona intestinalis	XP_002129392.1
Strongylocentrotus purpuratus	NP_001035495.1
Saccoglossus kowalevskii	XP_006813223.1
Capitella telata	ELU10181.1
Crassostrea gigas	EKC32876.1
Drosophila melanogaster	AAF36967.1
Caenorhabditis elegans	NP_498900.1
Hydra vulgaris	XM_004209432.1
Trichoplax adhaerens	XP_002109268.1
Monosiga brevicollis	MONBRDRAFT_39276.2
Salpingoeca rosetta	PTSG_04091.1

Table A1. Pins homologs with sequences shown in Figure 9b and Figure 14b, including species names and accession numbers.

Species	Phylum	GK Enzyme Accession #
Salpingoeca rosetta	Filozoa, Choanoflagelleta	PTSG_01650
Monosiga brevicollis	Filozoa, Choanoflagelleta	MONBRDRAFT_33796
Capsaspora owczarzaki	Filozoa, Filasterea	CAOG_06482

Table A2. List of orthologs of filozoan guanylate kinase enzyme and their gene accession numbers. Each gene is a reciprocal best hit of a human or yeast guanylate kinase query sequence.

REFERENCES CITED

1. Pawson T, Nash P.2000. Protein-protein interactions define specificity in signal transduction. Genes Dev. May 1;14(9):1027-47.

2. Scaffold proteins: hubs for controlling the flow of cellular information. 2011.Good MC, Zalatan JG, Lim WA. Science. May 6;332(6030):680-6. doi: 10.1126/science.1198701.

3. Nithianantharajah J, Komiyama NH, McKechanie A, Johnstone M, Blackwood DH, St Clair D, Emes RD, van de Lagemaat LN, Saksida LM, Bussey TJ, Grant SG. Synaptic scaffold evolution generated components of vertebrate cognitive complexity. Nat Neurosci. 2013 Jan;16(1):16-24. doi: 10.1038/nn.3276. Epub 2012 Dec 2.

4. Funke L, Dakoji S, Bredt DS. Membrane-associated guanylate kinases regulate adhesion and plasticity at cell junctions. Annu Rev Biochem 74, 219-245 (2005).

5. Oliva C, Escobedo P, Astorga C, Molina C, Sierralta J. 2012. Role of the MAGUK protein family in synapse formation and function. Dev Neurobiol. Jan;72(1):57-72. doi: 10.1002/dneu.20949.

6. Zheng CY, Seabold GK, Horak M, Petralia RS. 2011. MAGUKs, synaptic development, and synaptic plasticity. Neuroscientist. Oct;17(5):493-512. doi: 10.1177/1073858410386384.

7. Woods, D. F. & Bryant, P. J. (1991) The discs-large tumor suppressor gene of Drosophila encodes a guanylate kinase homolog localized at septate junctions, Cell 66, 4512464.

8. Li Y, Zhang Y, Yan H. Kinetic and thermodynamic characterizations of yeast guanylate kinase. Journal of Biological Chemistry (1996).

9. Kuhlendahl S, Spangenberg O, Konrad M, Kim E, Garner CC. Functional analysis of the guanylate kinase-like domain in the synapse-associated protein SAP97. Eur J Biochem. 1998 Mar 1;252(2):305-13.

10. SD Dimitratos, DF Woods, DG Stathakis, PJ Bryant. 1999. Signaling pathways are focused at specialized regions of the plasma membrane by scaffolding proteins of the MAGUK family. Bioessays 21, 912-921.

11. Olsen O, Bredt DS. Functional analysis of the nucleotide binding domain of membrane-associated guanylate kinases. J Biol Chem. 2003 Feb 28;278(9):6873-8. Epub 2002 Dec 12.
12. Johnston CA, Doe CQ, Prehoda KE. Structure of an enzyme-derived phosphoprotein recognition domain PLoS One. 2012;7(4):e36014. doi: 10.1371/journal.pone.0036014. Epub 2012 Apr 24.

13. Zhu J, Shang Y, Xia C, Wang W, Wen W, Zhang M. Guanylate kinase domains of the MAGUK family scaffold proteins as specific phospho-protein-binding modules. EMBO J. 2011 Nov 25;30(24):4986-97. doi: 10.1038/emboj.2011.428.

14. Enomoto M, Igaki T. Deciphering tumor-suppressor signaling in flies: genetic link between Scribble/Dlg/Lgl and the Hippo pathways.J Genet Genomics. 2011 Oct 20;38(10):461-70. doi: 10.1016/j.jgg.2011.09.005.

15. Chen K, Featherstone DE. Discs-large (DLG) is clustered by presynaptic innervation and regulates postsynaptic glutamate receptor subunit composition in Drosophila.BMC Biol. 2005 Jan 8;3:1.

16. Lu MS, Johnston CA. Molecular pathways regulating mitotic spindle orientation in animal cells. Development 140, 1843-1856 (2013).

17. Johnston CA, Hirono K, Prehoda KE, Doe CQ. Identification of an Aurora-A/PinsLINKER/Dlg spindle orientation pathway using induced cell polarity in S2 cells. Cell 138, 1150-1163 (2009).

18. Gillies TE, Cabernard C. Cell division orientation in animals. Curr Biol 21, R599-609 (2011).

19. Cabernard C, Doe CQ. Apical/basal spindle orientation is required for neuroblast homeostasis and neuronal differentiation in Drosophila. Dev Cell 17, 134-141 (2009).

20. Williams SE, Beronja S, Pasolli HA, Fuchs E. Asymmetric cell divisions promote Notch-dependent epidermal differentiation. Nature 470, 353-358 (2011).

21. Siegrist SE, Doe CQ. Microtubule-induced Pins/Galphai cortical polarity in Drosophila neuroblasts. Cell 123, 1323-1335 (2005).

22. Johnston CA, Whitney DS, Volkman BF, Doe CQ, Prehoda KE. Conversion of the enzyme guanylate kinase into a mitotic-spindle orienting protein by a single mutation that inhibits GMP-induced closing. Proc Natl Acad Sci USA 108, E973-8 (2011).

23. de Mendoza A, Suga H, Ruiz-Trillo I. Evolution of the MAGUK protein gene family in premetazoan lineages. BMC Evol Biol 10, 93 (2010).

24. te Velthuis AJ, Admiraal JF, Bagowski CP. Molecular evolution of the MAGUK family in metazoan genomes. BMC Evol Biol 7, 129 (2007).

25. Harms MJ, Thornton JW. Analyzing protein structure and function using ancestral gene reconstruction. Curr Opin Struct Biol 20, 360-366 (2010).

26. Harms MJ, Thornton JW. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. Nat Rev Genet 14, 559-571 (2013).

27. Grosberg RK, Strathmann RR. The evolution of multicellularity: a minor major transition? Annual Review of Ecology 38, 621-654 (2007).

28. Maynard Smith J, Szathmary E. The major transitions in evolution (W.H. Freeman Spektrum, Oxford ; New York, 1995).

29. Morin X, Bellaiche Y. Mitotic spindle orientation in asymmetric and symmetric cell divisions during animal development. Dev Cell 21, 102-119 (2011).

30. Bonner JT. The origins of multicellularity. Integrative Biology 1, 27-36 (1998).

31. King N. The unicellular ancestry of animal development. Dev Cell 7, 313-325 (2004).

32. King N, Hittinger CT, Carroll SB. Evolution of key cell signaling and adhesion protein families predates animal origins. Science 301, 361-363 (2003).

33. Nichols SA, Dirks W, Pearse JS, King N. Early evolution of animal cell signaling and adhesion genes. Proc Natl Acad Sci U S A 103, 12451-12456 (2006).

34. Richter DJ, King N. The genomic and cellular foundations of animal origins. Annu Rev Genet 47, 509-537 (2013).

35. Rokas A. The origins of multicellularity and the early history of the genetic toolkit for animal development. Annu Rev Genet 42, 235-251 (2008).

36. Nakajima Y, Meyer EJ, Kroesen A, McKinney SA, Gibson MC. Epithelial junctions maintain tissue architecture by directing planar spindle orientation. Nature 500, 359-362 (2013).

37. Cabernard C, Doe CQ. Apical/basal spindle orientation is required for neuroblast homeostasis and neuronal differentiation in Drosophila. Dev Cell 17, 134-141 (2009).

38. Williams SE, Beronja S, Pasolli HA, Fuchs E. Asymmetric cell divisions promote Notch-dependent epidermal differentiation. Nature 470, 353-358 (2011).

39. SE, Doe CQ. Microtubule-induced Pins/Galphai cortical polarity in Drosophila neuroblasts. Cell 123, 1323-1335 (2005).

40. Bilder D, Li M, Perrimon N. Cooperative regulation of cell polarity and growth by Drosophila tumor suppressors. Science 289, 113-116 (2000).
41. Bergstralh DT, Lovegrove HE, St Johnston D. Discs large links spindle orientation to apical-basal polarity in Drosophila epithelia. Curr Biol 23, 1707-1712 (2013).

42. Woods DF, Hough C, Peel D, Callaini G, Bryant PJ. Dlg protein is required for junction structure, cell polarity, and proliferation control in Drosophila epithelia. J Cell Biol 134, 1469-1482 (1996).

43. Dayel MJ, Alegado RA et al. Cell differentiation and morphogenesis in the colony-forming choanoflagellate Salpingoeca rosetta. Dev Biol 357, 73-82 (2011).

44. Blaszczyk J, Li Y, Yan H, Ji X. Crystal structure of unligated guanylate kinase from yeast reveals GMP-induced conformational changes. J Mol Biol 307, 247-257 (2001).

45. Sekulic N, Shuvalova L, Spangenberg O, Konrad M, Lavie A. Structural characterization of the closed conformation of mouse guanylate kinase. J Biol Chem 277, 30236-30243 (2002).

46. Harms MJ, Eick GN et al. Biophysical mechanisms for large-effect mutations in the evolution of steroid hormone receptors. Proc Natl Acad Sci U S A 110, 11475-11480 (2013).

47. Tokuriki N, Tawfik DS. Protein dynamism and evolvability. Science 324, 203-207 (2009).

48. Zhu J, Shang Y, Xia C, Wang W, Wen W, Zhang M. Guanylate kinase domains of the MAGUK family scaffold proteins as specific phospho-protein-binding modules EMBO J. 2011 Nov 25;30(24):4986-97. doi: 10.1038/emboj.2011.428.

49. Ferrari S, Marin O, Pagano MA, Meggio F, Hess D, El-Shemerly M, Krystyniak A, Pinna LA. Aurora-A site specificity: a study with synthetic peptide substrates. Biochem J. 2005 Aug 15;390(Pt 1):293-302.

50. Lye MF, Fanning AS, Su Y, Anderson JM, Lavie A. Insights into regulated ligand binding sites from the structure of ZO-1 Src homology 3-guanylate kinase module.J Biol Chem. 2010 Apr 30;285(18):13907-17.

51. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797 (2004).

52. Schultz J, Milpetz F, Bork P, Ponting CP. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* **95**, 5857-5864 (1998).

53. Guindon S, Dufayard JF et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321 (2010).

54. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* **60**, 685-699 (2011).

55. Yang Z, Kumar S, Nei M. A new method of inference of ancestral nucleotide and amino acid sequences. *Genetics* **141**, 1641-1650 (1995).

56. Hanson-Smith V, Kolaczkowski B, Thornton JW. Robustness of ancestral sequence reconstruction to phylogenetic uncertainty. *Mol Biol Evol* **27**, 1988-1999 (2010).

57. Agarwal KC, Miech RP, Parks REJ. Guanylate kinases from human erythrocytes, hog brain, and rat liver. *Methods Enzymol* **51**, 483-490 (1978).

58. King N, Young SL, Abedin M, Carr M. Starting and maintaining Monosiga brevicollis cultures. Cold Spring Harb Protoc. 2009 Feb.