

Article

# Domain Analysis and Motif Matcher (DAMM): A program to predict selectivity determinants in *Monosiga brevicollis* PDZ domains using human PDZ data

Haley A. Wofford<sup>1†</sup>, Josh Myers-Dean<sup>2†</sup>, Brandon A. Vogel<sup>1</sup>, Kevin Alexander Estrada Alamo<sup>1</sup>, Frederick A. Longshore-Neate<sup>1</sup>, Filip Jagodzinski<sup>2</sup>, and Jeanine F. Amacher<sup>1,\*</sup>

<sup>1</sup> Department of Chemistry, Western Washington University, Bellingham, WA, USA; email addresses: [haley.wofford@gmail.com](mailto:haley.wofford@gmail.com), [vogelb@wwu.edu](mailto:vogelb@wwu.edu), [estrack@wwu.edu](mailto:estrack@wwu.edu), [longshf@wwu.edu](mailto:longshf@wwu.edu)

<sup>2</sup> Department of Computer Science, Western Washington University, Bellingham, WA; email addresses: [myersdj@wwu.edu](mailto:myersdj@wwu.edu), [jagodz@wwu.edu](mailto:jagodz@wwu.edu)

<sup>†</sup> These authors contributed equally to this work.

\* Correspondence: Email: [amachej@wwu.edu](mailto:amachej@wwu.edu); Tel.: (360) 650-4397

**Abstract:** Choanoflagellates are single-celled eukaryotes with complex signaling pathways. They are considered the closest non-metazoan ancestors to mammals and other metazoans, and form multicellular-like states called *rosettes*. The choanoflagellate *Monosiga brevicollis* contains over 150 PDZ domains, an important peptide-binding domain in all three domains of life (Archaea, Bacteria, and Eukarya). Therefore, an understanding of PDZ domain signaling pathways in choanoflagellates may provide insight into the origins of multicellularity. PDZ domains recognize the C-terminus of target proteins and regulate signaling and trafficking pathways, as well as cellular adhesion. Here, we developed a computational program, Domain Analysis and Motif Matcher (DAMM), that predicts target specificity in choanoflagellate PDZ domains by analyzing peptide-binding cleft sequence identity as compared to human PDZ domains. We used this program, protein biochemistry, fluorescence polarization, and structural analyses to characterize the specificity of A9UPE9\_MONBE, a *M. brevicollis* PDZ domain-containing protein with no homology to any metazoan protein, finding that its PDZ domain is most similar to those of the DLG family. We then identified two endogenous sequences that bind A9UPE9 PDZ with <100  $\mu$ M affinity, a value commonly considered the threshold for cellular PDZ-peptide interactions. Taken together, this approach can be used to predict cellular targets of previously uncharacterized PDZ domains in choanoflagellates and other organisms. Our data contributes to investigations into choanoflagellate signaling and how it informs metazoan evolution.

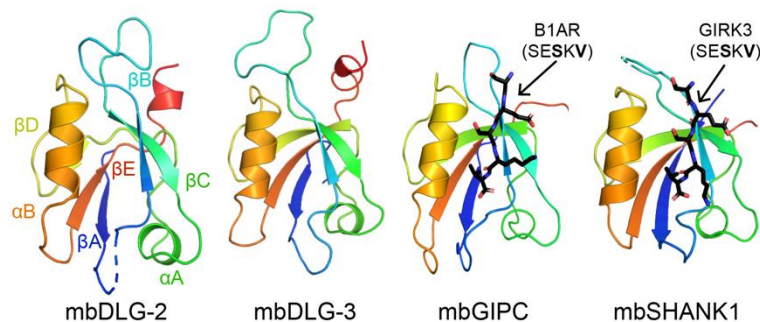
**Keywords:** protein-protein interactions, PDZ domains, choanoflagellates, evolution, target selectivity, protein-peptide interactions, signaling

## 1. Introduction

PDZ domains are small peptide-binding domains named for the proteins where they were first discovered; PSD-95, Dlg1, and ZO-1 (PDZ) [1–4]. These are scaffolding domains that recognize the extreme C-terminus of target proteins, with the bulk of the PDZ-protein interaction involving up to 6 residues on the target [5–7]. PDZ domains are found in all 3 domains of life (Archaea, Bacteria, and Eukarya) and PDZ-mediated interactions are important in signaling and trafficking pathways, as well as in cell adhesion [5,8,9]. Examples include: regulation of the trafficking of receptors throughout the cell, e.g., the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR), signaling pathways mediated by G-protein Coupled Receptors (GPCRs), and interactions with proteins involved in tight junctions and within the postsynaptic densities of neurons [10–27]. Due to their central role in cellular processes, PDZ domains are therapeutic targets in a number of human diseases [28–32]. They are also a common target of viral proteins during

infection. The best studied example includes PDZ domain targeting by oncoproteins in human papillomaviruses, but several other viruses, including influenza, human immunodeficiency virus (HIV), and coronaviruses also target PDZ domains [5,33–38]. For example, proteins in the SARS-CoV-2 proteome, which causes COVID-19 disease, interact with and bind several PDZ domains during infection [39–42]. The trafficking of angiotensin-converting enzyme 2 (ACE2) receptor, which is targeted by SARS-CoV-2, contains a PDZ-binding motif at its C-terminus (sequence: DVQISE, motif residues are underlined) and is also regulated by PDZ domains [43].

There are approximately 270 PDZ domains in the human proteome and PDZ domain-containing proteins have 1-13 PDZ domains [5]. All PDZ domains characterized to date share a common structural fold, consisting of an antiparallel  $\beta$ -sheet ( $\beta$ A- $\beta$ E) and 1-2  $\alpha$ -helices ( $\alpha$ A- $\alpha$ B) (**Figure 1**) [5]. Historically, PDZ domains are separated into binding classes, dependent on two positions, the most C-terminal residue, termed  $P^0$ , and two adjacent, or  $P^{-2}$  [5,44]. For example, Class I PDZ domains recognize the motif X-S/T-X- $\Phi$ , where X=any amino acid and  $\Phi$ =hydrophobic amino acid (typically, F/I/L/V) [5–7,44]. However, these motifs are unable to describe the overlapping specificities of the PDZ family [5]. Over the past decade, ourselves and others have teased apart selectivity determinants at positions other than those described in the characterized motifs [5,45–47]. The ability to identify and determine selectivity determinants in closely-related PDZ domains is critical to understanding their role in cellular pathways and processes.



**Figure 1.** Structures of PDZ domains from *Monosiga brevicollis* are shown in cartoon representation, with the conserved secondary structure elements labeled in the mbDLG-2 figure. Bound peptides are in black sticks and colored by heteroatom (N=blue, O=red), and labeled. PDB ID codes for choanoflagellate PDZ domains from this work include: 6X1X, 6X20, 6X22, 6X23, 6X1P, 6X1N, and 6X1R [48].

Choanoflagellates are microbial eukaryotes that live in marine and freshwater environments [49]. They are considered the closest relative to metazoans, and evolutionary biologists argue that these organisms can provide insight into the origins of multicellularity [49,50]. For example, choanoflagellates have extensive signaling networks and form a molecular architecture that resembles a multicellular state, called *rosettes* [51–53]. In our previous work, we concluded that there are 178 PDZ domains in *M. brevicollis*, with 1-20 PDZ domains in a single protein [48]. We previously used protein biochemistry and structural biology to characterize several PDZ domains from the *M. brevicollis* proteome, including those from mbDLG, mbGIPC, and mbSHANK1 (**Figure 1**) [48,54]. These PDZ domain-containing proteins all have clear homologues in the human proteome and we found that selectivity determinants are largely conserved, despite ~750 million years of evolution between these organisms [48].

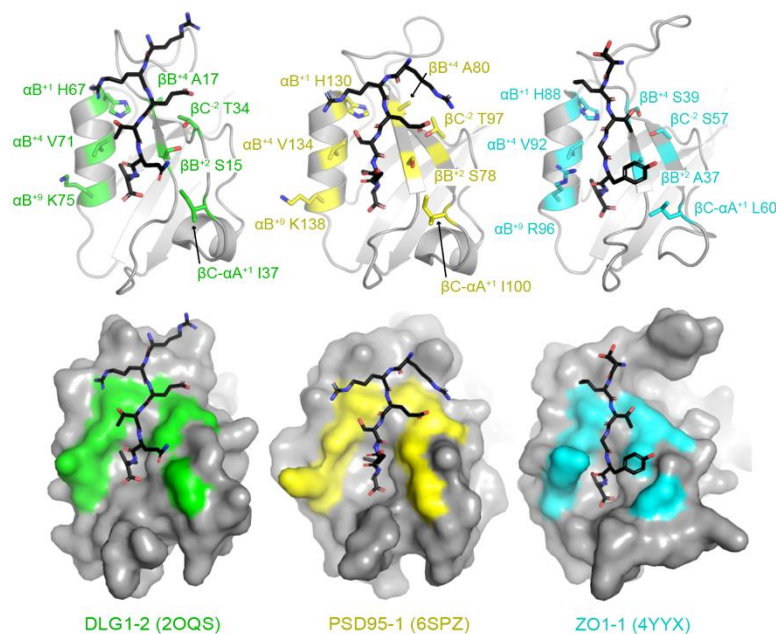
In this work, we wanted to expand the investigation of choanoflagellate PDZ domains and develop techniques to characterize those that do not have homologues in the human proteome. Based on our previous work and others, we hypothesized that binding preferences in PDZ domains are directly determined by the amino acids that interact with specific peptide positions, at both motif and modulator (or non-motif) positions [45–47,55–58]. Because we aimed to identify endogenous interactions, we chose not to use common high throughput techniques to determine PDZ selectivity, e.g., phage display or peptide arrays, which are tuned to look for highest affinity binders [32,38,47,59,60]. We reasoned

that this approach may limit the number of matching sequences in the *M. brevicollis* proteome and cause us to overlook sequences with lower affinity. It is well established that endogenous PDZ interactions are not necessarily optimized for affinity, as these are often regulatory in nature and in pathways involving several proteins and molecular states [5]. Therefore, in our approach, we developed a computational program, Domain Analysis and Motif Matcher (DAMM), which determines the human PDZ domain(s) that share the greatest number of conserved residues in the peptide-binding cleft with an input PDZ sequence, independent of overall sequence identity. The target sequences of human PDZ domains are largely known; therefore, we then predicted potential target sequences in the choanoflagellate proteome using the Motif Matcher part of the DAMM, based on our published *MotifAnalyzer-PDZ* program [54]. We used protein biochemistry, structural biology, and DAMM to identify two *M. brevicollis* sequences that bind to a previously uncharacterized PDZ domain with <100  $\mu\text{M}$  affinity validating our approach. We argue that this computational and experimental pipeline provides a straightforward methodology for predicting cellular targets of PDZ domains and can be applied to other peptide-binding domains, e.g., SH2 and SH3 domains, that also contain conserved structural folds.

## 2. Results

### 2.1. Identification of peptide-interacting residues in PDZ domains

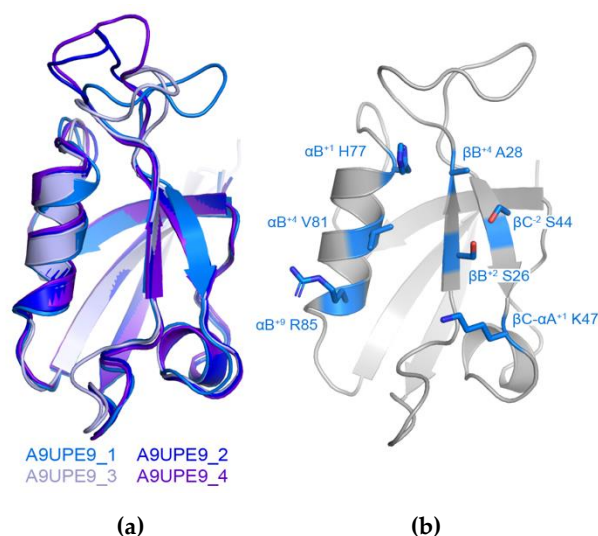
We wanted to develop a computational pipeline that could predict the endogenous targets of a PDZ domain that does not have a homologue in *Homo sapiens*. Based on our previous work characterizing PDZ domains in the *Monosiga brevicollis* proteome, we chose to focus on UniProt ID A9UPE9\_MONBE, a 545 amino acid protein with one PDZ and one SH3 domain. We will refer to this protein as A9UPE9 moving forward for simplicity. Notably, there are no proteins in the human proteome that contain this domain architecture and a BlastP search of the full-length sequence returns alignments limited to one of the isolated domains.



**Figure 2.** Identification of peptide-binding cleft residues in DLG1 PDZ2 (or DLG1-2, PDB ID 2OQS, green), PSD95-1 (6SPZ, yellow), and ZO1-1 (4YYX, cyan). Proteins are shown in both cartoon (top) and surface representations (bottom) with bound peptides in black sticks and colored by heteroatom (N=blue, O=red). The peptide-binding cleft residues used in this analysis are represented as side chain sticks, and colored by heteroatom or monochrome, respectively, and labeled. DLG1-2 is bound to a peptide of the HPV18 E6 oncoprotein (sequence: RRETQV) [61]. PSD95-1 is bound to a peptide with sequence RRESEI [62]. ZO1-1 is bound to the Claudin2 C-terminal tail (SLTGYV) [63].

Using structural analyses of peptide-bound PDZ domains, we identified 7 key residues whose side chains interact directly with the peptide. Examples include peptide-bound structures of DLG1 PDZ2 (or DLG1-2), PSD95-1 (which is also referred to as DLG4-1), and ZO1-1 (**Figure 2**) [61–63]. These include: the  $\beta B^{+2}$ ,  $\beta B^{+4}$ ,  $\beta C^{-2}$ ,  $\beta C^{-\alpha A^{+1}}$ ,  $\alpha B^{+1}$ ,  $\alpha B^{+5}$ , and  $\alpha B^{+9}$  residues, where the superscript numbers indicate residues from the N- (positive values) or C-termini (negative values) of those secondary structure elements, and “ $\beta C^{-\alpha A}$ ” is the loop between the  $\beta C$  strand and  $\alpha A$  helix. Notably, due to a large degree of variability in the length of the  $\beta B$ - $\beta C$  loop amongst PDZ domains, we did not include residues in this structural element, although they can interact with peptide positions, e.g., P<sup>-4</sup> and P<sup>-5</sup>. In addition, other residues make non-covalent interactions with peptide residues in individual PDZ domains, but we kept our assignments relatively conservative, limited to those that form the core of the peptide-binding cleft (**Figure 2**).

We also wanted to use the online Robetta server for *de novo* structure prediction to generate a model of A9UPE9 (**Figure 3**) [64]. Robetta output 4 conformations that mainly differed in the conformation of the  $\beta B$ - $\beta C$  loop and alignments revealed RMSD values for main chain atoms, as compared to conformation 1 of: 0.455 Å for 289 atoms (conformation 2), 0.530 Å for 291 atoms (conformation 3), and 0.679 Å for 311 atoms (conformation 4) (**Figure 3a**). Due to the similarities, we restricted our structural analyses to conformation 1. The structure, as well as sequence patterns, allowed us to identify the relevant peptide-binding cleft residues (**Figure 3b**).



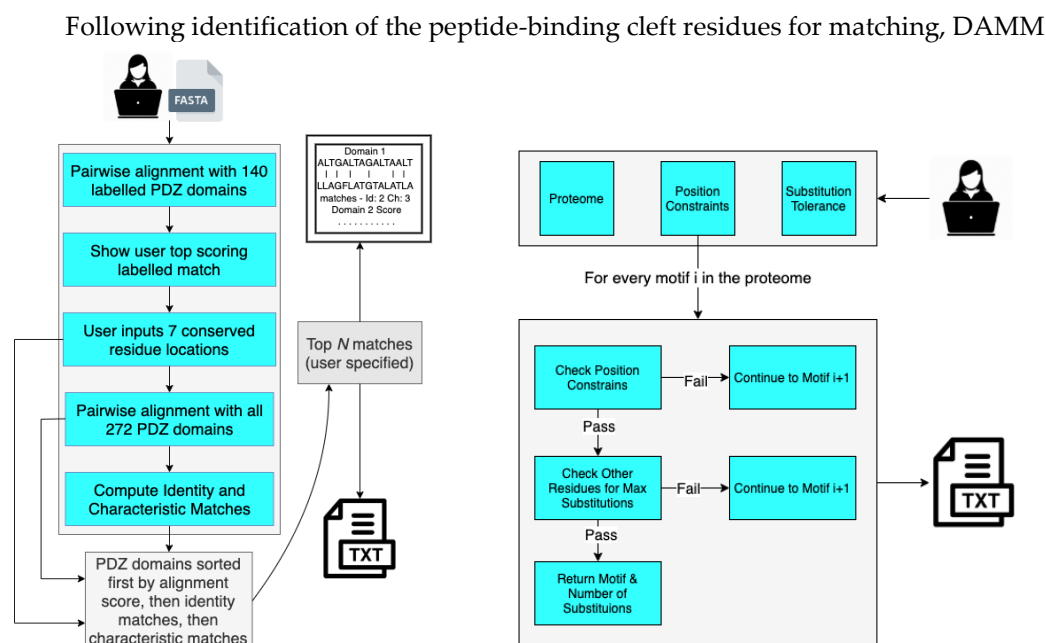
**Figure 3.** Model of A9UPE9 PDZ structure. The 4 conformations from Robetta are in cartoon representation and colored as labeled in (a). (b) Identification of peptide-binding cleft residues in A9UPE9 conformation 1 (A9UPE9\_1). Residues used for analysis are shown as side chain sticks, colored marine by heteroatom (N=blue, O=red), and labeled.

## 2.2. Domain Analysis and Motif Matcher (DAMM) program

In order to determine the human PDZ domain with the most similar peptide-binding cleft as A9UPE9, we developed a Python-based program called Domain Analysis and Motif Matcher (DAMM). The computational pipeline of DAMM is in **Figure 4a**. We will refer to the input sequence as “A9UPE9” or “the choanoflagellate PDZ domain sequence,” but this program can theoretically be used for any PDZ domain sequence from any organism. Ultimately, we did not want to require a structural model of an input PDZ domain sequence in order to identify the 7 peptide-binding cleft residues. Therefore, we manually identified these amino acids for 140 solved PDZ domain structures, based on our previous curation of mammalian PDZ domains in the Protein Data Bank [5]. This data became an input database for our computational program and the first step in DAMM is to take a single input sequence in FASTA file format and run a BLAST alignment with the 140 PDB sequences that we manually curated [5,65]. The top pairwise alignment based on sequence

similarity is provided as an output with asterisks marking the peptide-binding cleft residues of the annotated PDB structure. The user input sequence has reference numbers, allowing the user to identify and input the 7 peptide-binding cleft positions for matching. For example, for A9UPE9, if the input sequence was of the full-length protein, the user would input: 26, 28, 44, 47, 77, 81, and 85 (**Figure 3b**). It is important to note that these numbers may vary depending on the input sequence, e.g., if the input FASTA file only contained the UniProt annotated A9UPE9 PDZ domain (residues 12-99), the numbers would be shifted by 12 as a result because E12 would be considered residue 1 by DAMM.

**Figure 4.** The Domain Analysis and Motif Matcher (DAMM) program. Details of the computational workflow of the Domain Analysis part of the program are in (a) and Motif Matcher piece are in (b).



runs BLAST pairwise sequence alignments with all previously curated 272 human PDZ domain sequences, including those that do not have a solved structure in the Protein Data Bank [5]. The user is prompted to specify how many sequences to output, listed in descending order by overall sequence identity. Included in the output file is a list of the pairwise alignments with number of peptide-binding cleft matches, in both identity and similarity, indicated for each of the alignments. The output file is ordered by sequence identity to ensure that alignments listed at the top are those where the majority of the PDZ domain sequences aligned, but overall, sequence identity does not play a role in number of matches specified. Amino acids are defined as “similar” based on the following categories, using one-letter abbreviations: acidic (D, E), basic (H, K, R), polar (C, M, N, Q, S, T), non-polar (A, I, L, V), and aromatic (F, W, Y).

Once these results are obtained, the user can run literature and database searches to identify potential binding sequences, based on known human PDZ targets. The Motif Matcher piece of DAMM can then be used similar to *MotifMatcher-PDZ*, where C-terminal sequences in the choanoflagellate proteome are filtered and screened for closest matches to an input sequence [54]. Alternatively, we can use DAMM to screen the *M. brevicollis* proteome for specific preferences at each position, similar to those that would be identified using phage display or another high throughput technique. All components of DAMM were developed in Python 3, and are available at <https://github.com/joshmyersdean/damm>.

### 2.3. Binding preferences of A9UPE9 PDZ

To identify the binding preferences of A9UPE9 PDZ, we used the DAMM program to find the human PDZ domains with the most similar binding cleft residues. For reference, the PDZ domains with the top 5 highest overall sequence identities are in **Table 1**.

**Table 1.** Highest sequence identity values of human PDZ domains with A9UPE9 PDZ.

PDZ Domain	Sequence Identity	Sequence Similarity
INADL-10	42% (35/83)	59% (49/83)
SNTG2	41% (36/87)	57% (50/87)
GRIP1-2	41% (30/74)	60% (45/74)
GRIP2-2	41% (26/63)	61% (39/63)
GRIP1-4	35% (30/85)	51% (44/85)

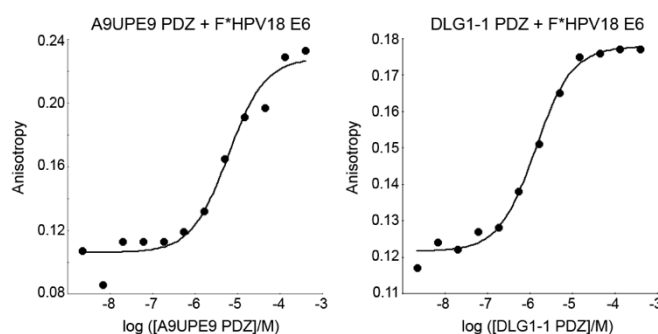
Based on the overall sequences, A9UPE9 appears to most closely resemble the GRIP PDZ domains (**Table 1**). However, the results from the DAMM program revealed a different result. **Table 2** shows the human PDZ domains with the most similar peptide-binding clefts as A9UPE9 PDZ. For reference, the GRIP1-2 and A9UPE9 PDZ domains have 1 identical residue amongst the peptide-binding cleft positions, an  $\alpha B^{+1}$  His, and 3 similar residues, at the  $\beta C^{-2}$  (Ser for A9UPE9, Thr for GRIP1-2),  $\beta C-\alpha A^{+1}$  (Lys and Arg, respectively), and  $\alpha B^{+9}$  (Arg and Lys, respectively). This result highlights the importance of using number of matches, and not sequence identity, for this type of analysis.

**Table 2.** DAMM program results for A9UPE9 PDZ.

PDZ Domain	Matching/Similar Residues	Sequence Identity
SYNJ2BP	5/2	31%
DLG1-2	4/2	29%
DLG2-2	4/2	29%
DLG4-2	4/2	29%
MPDZ-12	4/2	32%
DLG3-2	4/2	30%
DLG1-1	4/2	30%
DLG2-1	4/2	29%
DLG4-1	4/2	29%
DLG3-1	4/2	30%

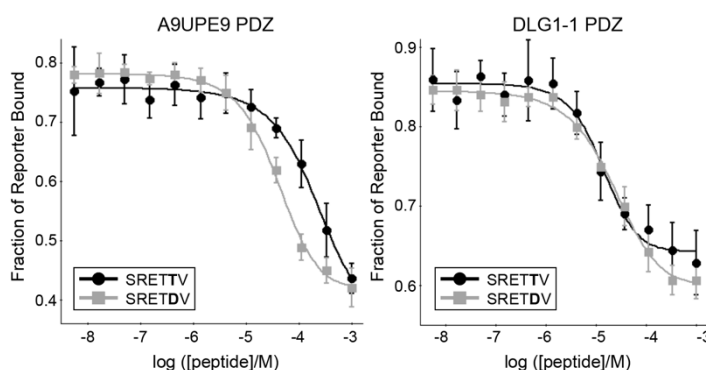
Our DAMM program results clearly show that while the top result for A9UPE9 PDZ is SYNJ2BP PDZ, overall its peptide-binding cleft is most similar to the DLG family of PDZ domains. Of these PDZ domains, those from DLG proteins are more extensively studied, including by our lab. Therefore, we chose to use DLG1-1 PDZ in fluorescence polarization experiments to directly compare binding affinity values with A9UPE9 PDZ.

The A9UPE9 and DLG1-1 PDZ domains were expressed and purified following similar protocols as previously described and in the Materials and Methods [48,54]. The reporter peptide used for both proteins was a fluoresceinated sequence matching the C-terminus of the HPV18 E6 oncoprotein (sequence: FITC-RLQRRRETQV), or  $F^*$ -HPV18 E6. Determined  $K_D$  values for each protein following triplicate experiments were:  $2.3 \pm 1.6 \mu M$  for DLG1-1 PDZ and  $6.9 \pm 1.7 \mu M$  for A9UPE9 PDZ (**Figure 5**). The HPV18 E6 sequence was chosen based on its known interaction with DLG1-1 PDZ [66–68]. A fluoresceinated version of the HPV16 E6 C-terminal sequence (FITC-SSRTRRETQL) was also tested, and bound A9UPE9 PDZ with  $K_D = 19.5 \pm 1.8 \mu M$ , based on duplicate experiments. Therefore, we used the higher affinity  $F^*$ -HPV18 E6 as our reporter peptide.



**Figure 5.** Representative fluorescence polarization experiments to determine  $K_D$  values of A9UPE9 (left, as labeled) and DLG1-1 (right, as labeled) PDZ domains. Overall  $K_D$  values were calculated from triplicate experiments.

We next wanted to interrogate preferences at specific positions along the peptide, as well as broadly compare binding affinities of known PDZ targets to the A9UPE9 and DLG1-1 PDZ domains. Starting with the HPV18 E6 C-terminal sequence, we designed a number of peptides to test specific peptide positions, including: HPV16 E6, HPV18 E6, SRETTV, SRETDV, RRETTV, and RRETDV, where underlined residues differ from HPV18 E6. As a control, we also tested human SNX27, SHANK1, and GIPC PDZ targets previously studied, including: GIRK3 (sequence: ESESKV), BPIX (WDETNL), GAIP (QSS-SEA), TYRP1 (PNQSVV), and B1AR (ASESKV), as well as previously identified mbSHANK targets, including: A9V7Z4 (EDETAL), A9UP44 (QSESRL), and A9UXE1 (QDETAL) [48,54]. Example competition experiment binding curves are in **Figure 6** and all  $K_i$  binding affinity values are reported in **Table 3**.



**Figure 6.** The binding isotherms from competition fluorescence polarization experiments of A9UPE9 (left, as labeled) and DLG1-1 (right, as labeled) PDZ domains are shown with the HPV18 E6 variant peptides: SRETTV and SRETDV. The average values are shown, with error bars representing standard deviation from at least triplicate experiments.  $K_i$  values are determined using SOLVER, as previously described [12,45,46,48,54,60].

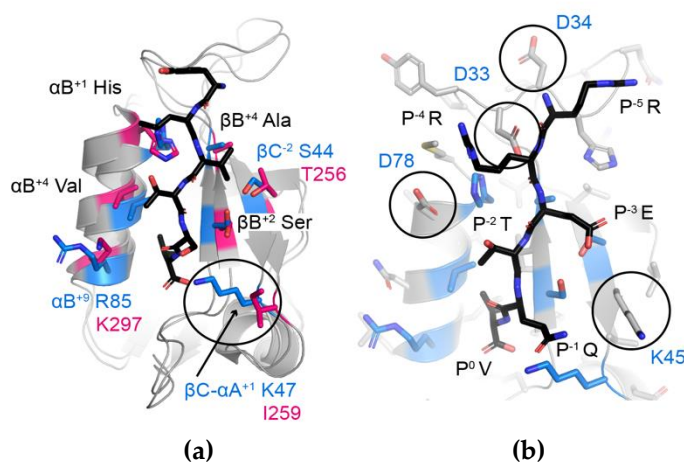
**Table 3.** Binding affinities for the A9UPE9 and DLG1-1 PDZ domains, determined by fluorescence polarization experiments. For all,  $F^*$ -HPV18 E6 was used as a reporter peptide and values reflect at least triplicate experiments. Differences as compared to HPV18 E6 are underlined, as relevant. Lines indicate related clusters of peptides: HPV sequences, HPV18 E6 variant sequences, SNX27/SHANK1/GIPC targets, and mbSHANK *M. brevicollis* targets, respectively.

Peptide	Sequence	$K_i$ values ( $\mu\text{M}$ )	
		A9UPE9 PDZ	DLG1-1 PDZ
HPV18 E6	RLQRRRETQV	$16 \pm 1$	$2.8 \pm 1.3$
HPV16 E6	SSRTRRETQL	$66 \pm 24$	$23 \pm 5$
<u>S</u> RETTV	<u>S</u> RETTV	$83 \pm 31$	$4.4 \pm 1.9$
<u>S</u> RETDV	<u>S</u> RETDV	$12 \pm 3$	$7.9 \pm 3.7$
RRETTV	RRETTV	$33 \pm 1$	$2.0 \pm 1.6$
RRETDV	RRETDV	$8.1 \pm 3.1$	$9.5 \pm 6.6$

GIRK3	EESKVV	>1000	350 ± 20
BPIX	WDETNL	330 ± 160	460 ± 140
GAIP	QSSSEA	>1000	980 ± 180
TYRP1	PNQSVV	97 ± 24	46 ± 21
B1AR	ASESKV	95 ± 30	44 ± 23
A9V7Z4	EDETAL	580 ± 160	460 ± 190
A9UP44	QSESRL	64 ± 35	39 ± 25
A9UXE1	QDETAL	460 ± 210	190 ± 20

As expected, both A9UPE9 and DLG1-1 bound the HPV18 E6 variant peptides with much higher affinity than peptides matching SNX27, SHANK1, and GIPC targets (**Table 3**). Our competition experiments revealed a key difference between the A9UPE9 and DLG1-1 PDZ domains. Specifically, while DLG1-1 PDZ bound the HPV18 E6 variant peptides containing a P<sup>-1</sup> Thr residue with higher affinity, A9UPE9 PDZ greatly preferred a P<sup>-1</sup> Asp (**Table 3**). Sequence and structural analyses suggest this difference is due to the  $\beta$ C- $\alpha$ A<sup>+1</sup> residue, which is I259 in DLG1-1 and K47 in A9UPE9 PDZ, and which can directly interact with the P<sup>-1</sup> residue (**Figure 7a**). This  $\beta$ C- $\alpha$ A<sup>+1</sup> Ile is also in DLG1-2 and PSD95-1 (**Figure 2**). In fact, a multiple sequence alignment of the 12 PDZ domains in DLG1/2/3/4 confirm that these proteins share almost identical peptide-binding clefts, with  $\beta$ B<sup>+2</sup> = S or N,  $\beta$ B<sup>+4</sup> = A or V,  $\beta$ C<sup>-2</sup> = S or T,  $\beta$ C- $\alpha$ A<sup>+1</sup> = I or L,  $\alpha$ B<sup>+1</sup> = H,  $\alpha$ B<sup>+5</sup> = V or A, and  $\alpha$ B<sup>+9</sup> = K.

Finally, we predicted that the apparent preferences of A9UPE9 PDZ for P<sup>-4</sup> and P<sup>-5</sup> Arg residues (e.g., for P<sup>-5</sup>, in RRETQV as compared to SRETQV) are due to D33 and/or D34, residues in the  $\beta$ B- $\beta$ C loop, as well as D78, the amino acid following the  $\alpha$ B<sup>+1</sup> His, and that a preference for a P<sup>-3</sup> Glu (as in the HPV sequences) is likely due to K45 (**Figure 7b**). Taken together, our DAMM program successfully identified DLG PDZ domains as having similar binding preferences as A9UPE9 PDZ, based on shared binding cleft residues. A small number of substitutions in a target sequence, here HPV18 E6, were able to further refine these preferences for A9UPE9 PDZ and we identified two sequences, F\*-HPV18 E6 and RRETQV, that bound with <10  $\mu$ M affinity.



**Figure 7.** (a) Comparison of the peptide-binding clefts of A9UPE9 PDZ (marine) and DLG1-1 (pink, PDB ID 3RL7). DLG1-1 is bound to the peptide matching the C-terminal sequence of the Adenomatous Polyposis Coli (APC) tumor suppressor protein (sequence: YLVTSV) [69]. Peptide-binding cleft residues are shown as side chain sticks, colored by heteroatom (N=blue, O=red), and labeled, where black text indicates shared residues and blue/pink text indicates differences. The  $\beta$ C- $\alpha$ A<sup>+1</sup> residue, which influences P<sup>-1</sup> selectivity, is indicated with a black circle. (b) Additional residues in A9UPE9 that may influence peptide preferences include D33, D34, K45, and/or D78, which are labeled and highlighted with circles. All side chains are shown as gray sticks and colored by heteroatom, with peptide-binding cleft residues used in the DAMM program colored as in (a). The HPV18 E6 peptide (sequence: RRETQV) from PDB ID 2OQS is shown as reference and labeled.

#### 2.4. Endogenous targets of A9UPE9 PDZ.



Finally, we set out to identify potential endogenous targets of A9UPE9 PDZ, based on peptide-binding cleft similarities with human PDZ domains. We used our HPV18 E6 variant sequences (**Table 3**), as well as known human targets of the top hit from the DAMM program, SYNJ2BP PDZ. As previously described, Motif Matcher scans a filtered set of *M. brevicollis* sequences, consisting of the final 6 C-terminal residues for each protein in the UniProt-downloaded proteome for most-similar matches to each sequence [54]. We initially required the program to maintain a P<sup>0</sup> Val residue, based on the affinity difference between the HPV16 E6 and HPV18 E6 sequences, but allowed P<sup>-2</sup> to be Ser or Thr. In addition, we ran *M. brevicollis* proteins that were closest matches to known SYNJ2BP PDZ targets, including TMIGD1 (sequence: HSETAL), DLL1/DLL4 (VIATEV), and SYNJ2BP (ASGSSV), through Motif Matcher [70–72]. We ultimately tested 7 additional peptides in our fluorescence polarization assay, including A9V6G5\_MONBE (HRESTV), A9UWH1\_MONBE (STRSDV), A9UR52\_MONBE (SRRTEV), A9UWP5\_MONBE (GSESSV), and A9UYY4\_MONBE (RLASEV), as well as A9VB85\_MONBE (ARESEI), which contained P<sup>-1</sup>/P<sup>-3</sup> Glu residues and a P<sup>-4</sup> Arg, as well as the DLL1/DLL4 sequence directly (**Table 4**). A9VB85\_MONBE was identified by using Motif Matcher to search for sequences with the highest number of positive modulators for A9UPE9 based on our biochemical analyses, including P<sup>-1</sup> = E/D/Q/N, P<sup>-3</sup> = E/D, P<sup>-4</sup> = R/K, and P<sup>-5</sup> = any residue except E/D.

**Table 3.** Binding affinities for A9UPE9 PDZ with *M. brevicollis* sequences, determined by fluorescence polarization experiments. Experiments conducted as in **Table 3** and as described in the Materials and Methods.

Peptide	Sequence	K <sub>i</sub> (μM)
		A9UPE9 PDZ
A9V6G5_MONBE	HRESTV	97 ± 30
A9UWH1_MONBE	STRSDV	>1000
A9UR52_MONBE	SRRTEV	>1000
A9UWP5_MONBE	GSESSV	490 ± 110
A9UYY4_MONBE	RLASEV	>1000
A9VB85_MONBE	ARESEI	>1000
DLL1/DLL4	VIATEV	240 ± 40

As shown in **Table 4**, most of our peptides failed to bind A9UPE9 PDZ with reasonable affinities to suggest an endogenous interaction. Structural analyses suggested that A9UWH1\_MONBE and A9UR52\_MONBE likely failed to bind due to the presence of the P<sup>-3</sup> Arg residue, which may negatively interact with the positively-charged K45 (**Figure 7b**). It is not immediately clear why A9UYY4\_MONBE bound with no detectable affinity, defined as >1000 μM. One hypothesis is that A9UPE9 PDZ requires a P<sup>-3</sup> Glu for binding. Despite containing all other preferred selectivity determinants, including P<sup>-1</sup>/P<sup>-3</sup> Glu residues and a P<sup>-4</sup> Arg, the P<sup>0</sup> Ile in A9VB85\_MONBE likely explains why that sequence does not bind A9UPE9 PDZ. Our previous work investigating the basis of P<sup>0</sup> selectivity found that the PDZ-defining carboxylate-binding loop sequence “GFGF,” as in A9UPE9 PDZ, will likely only accommodate Val and Leu residues [46].

Overall, this work shows that it is challenging to predict endogenous targets of a choanoflagellate PDZ sequence that is not homologous to human PDZ domains. However, we identified one sequence, A9V6G5\_MONBE, that binds with an endogenously relevant affinity, <100 μM, using the DAMM program. In addition, a previously tested *M. brevicollis* peptide, A9UP44 (QSESRL), bound A9UPE9 PDZ with K<sub>i</sub> = 64 μM (**Table 3**). It is not clear why A9UPE9 PDZ bound this sequence with relatively high affinity, despite the P<sup>-1</sup> Arg residue, and suggests that the P<sup>-3</sup> Glu provides the majority of free energy of binding amongst non-motif/modulator residues. In our previous work, we successfully used *MotifMatcher*-PDZ to identify sequences that bound mbSHANK with ~40-70 μM affinity, which is in the same relative range and may reflect the relatively low affinities that PDZ domains have for their endogenous targets [5,54].

Taken together, structural and sequence analyses, in combination with the DAMM program, allowed us to identify two potential endogenous targets of A9UPE9 PDZ. Future work to validate these targets *in vivo*, as well as to test additional *M. brevicollis* sequences is necessary to develop a holistic view of the cellular function of A9UPE9.

### 3. Discussion

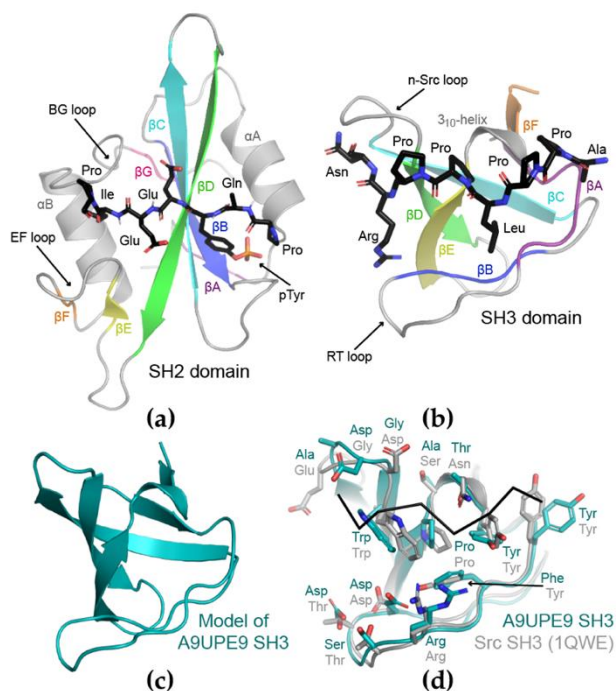
Characterization of the proteins in choanoflagellate signaling pathways can provide important insights into the evolution of human cellular processes. We developed a program to predict the target selectivity of PDZ domains in choanoflagellates, based on conserved amino acids in the peptide-binding cleft. Another critical consideration in future studies will be to think about the other domains in PDZ domain-containing proteins and how they coordinate to regulate cellular processes. For example, there are 8 proteins in the *M. brevicollis* proteome that contain both PDZ and SH3 domains, a domain architecture that is only present in choanoflagellates based on BLAST searches, including: A9UNP0\_MONBE, A9UPE9\_MONBE, A9UPI8\_MONBE, A9UUC6\_MONBE, A9UYE7\_MONBE, A9V111\_MONBE, A9V6P1\_MONBE, and A9V7E4\_MONBE. Notably, proteins that contain both PDZ and SH2 domains appear to be unique to organisms from the phylum *Choanozoa* as well and there are 5 *M. brevicollis* proteins that contain these domains: A9V6T4\_MONBE, A9V7X5\_MONBE, A9VA09\_MONBE, A9VB90\_MONBE, and A9VC25\_MONBE [48].

SH2 and SH3 domains were first discovered in the late 1980s, based on homology between the Src oncoprotein and phospholipase c, and play important scaffolding roles in signal transduction pathways [73–75]. There are >120 SH2 domains in the human proteome, and >300 SH3 domains [76,77]. SH2 domains recognize phosphotyrosine-containing peptide motifs, and share a common structural fold, characterized by a central 7-stranded  $\beta$ -sheet ( $\beta$ A- $\beta$ G) and 2  $\alpha$ -helices (**Figure 8a**) [78–81]. SH3 domains recognize proline-rich peptides that form a polyproline type II helix, and also share a common structural fold, characterized by a central 6-stranded  $\beta$ -sheet ( $\beta$ A- $\beta$ F) and a  $3_{10}$ -helix (**Figure 8b**) [82]. These domains play critical roles in tyrosine signaling pathways in mammalian cells, by coupling tyrosine phosphorylation to intracellular signaling [83–87].

Work over the past decade has identified a complex network of tyrosine kinase signaling in choanoflagellates, more so than in any metazoan characterized [88–92]. It is clear that such networks are critical for cellular communication, environmental adaptation, and other processes in both non-metazoans, such as choanoflagellates, as well as in metazoans [90]. However, the intricacies of metazoan signaling pathways are not well understood and future investigations into how PDZ and SH2/SH3 domains act synergistically in these proteins may provide deeper insight into the role of tyrosine kinase signaling in non-metazoans.

We developed the DAMM program, which can be used to predict endogenous target sequences of a previously uncharacterized PDZ domain, using sequence identity at key residues in the peptide-binding cleft. We hypothesize that the same approach could be applied to SH3 and other domains, based on conservation in amino acids that directly interact with a peptide target (**Figures 8c-d**). For example, despite only 29% sequence identity (14/48 residues), structural analyses of the A9UPE9 SH3 and human Src SH3 domains reveal similar peptide-binding cleft residues (**Figure 8d**). While protein biochemistry can tease apart unique selectivity determinants and identify potential protein-protein interactions in choanoflagellates, cell-based experiments to directly test

predictions and investigate these complex pathways will be an exciting next area of research in understanding the evolution of PDZ and tyrosine kinase signaling.



**Figure 8.** (a-b) Conserved structural folds of the Src SH2 (a), PDB ID 1SPS, and SrcSH3 (b), 1QWE, domains are shown as cartoon representation, with secondary structure elements labeled and  $\beta$ -strands colored. Bound peptides are in black sticks and colored by heteroatom (N=blue, O=red): for (a), sequence: PQPpYEEIP, and for (b), APP12 (APPLPPRN). (c) A homology model of A9UPE9 SH3 is shown in cartoon representation. The template PDB ID used was 7CSO (Ephexin4 SH3). (d) The A9UPE9 and Src SH3 domains are in cartoon representation, with peptide-binding cleft residues in sticks and labeled. The APP12 peptide is in black ribbon. Alignment of main chain atoms revealed an overall RMSD = 0.824 Å for 170 atoms.

#### 4. Materials and Methods

**Protein Expression and Purification.** Expression and purification of all human and *M. brevicollis* PDZ domains followed a similar protocol as previously reported for mbSHANK1 PDZ [48,54]. Briefly, N-terminal His-tagged versions of A9UPE9 PDZ (residues 7-98) and DLG1-1 PDZ (residues 220-317) with cleavable TEV sites were inserted into the pET28a+ vector by gene synthesis (GenScript) and expressed in *Escherichia coli* BL21 (DE3) cells. Cells were lysed on ice using sonication. The lysis buffer used was: 50 mM Tris pH 8.5, 200 mM NaCl, 10 mM CaCl<sub>2</sub>, 10 mM MgCl<sub>2</sub>, 20% (w/v) glycerol, 50 mM imidazole pH 8.5, 0.25 mM TCEP, DNase, and protease inhibitor cocktail. Immobilized metal-affinity chromatography, 5 mL HisTrap (GE Healthcare, now Cytiva), was used to purify proteins from the clarified supernatant. The wash buffer used was: 25 mM imidazole pH 8.5, 25 mM Tris pH 8.5, 250 mM NaCl, 10% (v/v) glycerol, and 0.25 mM TCEP, and elution buffer was: 400 mM imidazole pH 8.5, 25 mM Tris pH 8.5, 50 mM NaCl, 10% (v/v) glycerol, and 0.25 mM TCEP. Except for the protein used in binding experiments, the protein was then dialyzed in dialysis/gel filtration buffer (25 mM Tris pH 8.5, 150 mM NaCl, 10% (w/v) glycerol, 0.5 mM TCEP) and incubated at 4°C overnight with TEV protease to cleave off the His-tag. The cleaved protein was then purified using a second nickel column with the wash and elution buffers described above. All proteins were further purified on a Superdex S75 column (Cytiva), using gel filtration buffer. Proteins were concentrated using Amicon centrifugal concentrators (3 MWCO). Concentrated proteins used in fluorescence polarization assays were flash frozen in liquid nitrogen for storage at -80°C. The extinction coefficient values used for quantification of the proteins at A<sub>280</sub> were: 1490 cm<sup>-1</sup>\*M<sup>-1</sup> (A9UPE9) and 4470 cm<sup>-1</sup>\*M<sup>-1</sup> (DLG1-1).

*Fluorescence polarization.* Fluorescence polarization assays were conducted as previously described [12,45,46,48,54,60]. The reporter peptide used in competition experiments was F\*-HPV18 E6 (FITC-RLQRRRETQV) and  $K_D$  values were averaged from triplicate experiments. In both  $K_D$  and  $K_i$  experiments, 30 nM reporter peptide was used. Determined  $K_D$  values for each protein following triplicate experiments were:  $2.3 \pm 1.6 \mu\text{M}$  for DLG1-1 PDZ and  $6.9 \pm 1.7 \mu\text{M}$  for A9UPE9 PDZ. The final protein concentrations used for  $K_i$  competition experiments were: 10.5  $\mu\text{M}$  (A9UPE9) and 4.5  $\mu\text{M}$  (DLG1-1). Competition experiments were performed in at least triplicate and binding affinities were determined using SOLVER, as previously described [12,45,46,48,54,60]. Binding curves were visualized using Kaleidagraph.

*Design and development of Domain Analysis and Motif Matcher (DAMM).* The DAMM program was written using Python3. The first step of DAMM, Domain Analysis (DA), receives a sequence in FASTA format as input from the user. DA interfaces with BioPython and performs a pairwise BLAST-style alignment of the input sequence to the 140 PDZ labeled domains in our database [93]. DAMM outputs the top-scoring pairwise alignment, showing known conserved residues, then prompts the user for a list of 7 conserved residues.

Following identification of peptide-binding cleft residues, DA aligns the input sequence with the sequences of all 272 human PDZ domains, based on previous curation [5]. The output file lists pairwise alignments in order based on BlastP score, and also lists the number of matching and/or similar peptide-binding cleft residues. Amino acid similarity is based on the following 5 groupings: (group 1, hydrophobic) AILV; (group 2, negatively-charged) DE; (group 3, positively-charged) KRH; (group 4, aromatic) FYW; (group 5, polar) CMNQST. Glycine (G) and proline (P) are not included in the groupings because of their unique chemical and/or structural properties.

A user might want to perform a motif similarity search after the alignment analysis and determination of number of matching/similar peptide-binding cleft residues. DAMM provides a motif matching algorithm, inspired by *MotifAnalyzer-PDZ* and here named Motif Matcher (MM) [54]. Motif Matcher takes as input a target proteome, specific position constraints for each motif residue (e.g.,  $P^0 = \text{F, I, L, or V}$  and  $P^{-2} = \text{S or T}$ ), and the maximum amount of substitutions a user will tolerate (e.g., for a hexamer with 2 defined motif positions, the maximum number here would be equal to 4). The output is a list of all C-terminal hexameric sequences that meet these constraints, including the number of substitutions at non-motif positions (Table 4).

**Table 4.** Example output of Motif Matcher, using sequence GGGTGL as a search model.

Number of Substitutions	UniProt ID	Matching Sequence	Reference Sequence
2	tr A9UQN5 A9UQN5_MONBE	GGCTLL	GGGTGL
2	tr A9V457 A9V457_MONBE	YGGTSF	GGGTGL
2	tr A9UZY0 A9UZY0_MONBE	RYGSGV	GGGTGL
2	tr A9UWI6 A9UWI6_MONBE	GGCSLL	GGGTGL
2	tr A9UR91 A9UR91_MONBE	GYGSTI	GGGTGL
2	tr A9UU72 A9UU72_MONBE	GGPTDI	GGGTGL
2	tr A9UX10 A9UX10_MONBE	GLGTTI	GGGTGL
2	tr A9UNL7 A9UNL7_MONBE	GGSTQI	GGGTGL
2	tr A9UPF5 A9UPF5_MONBE	GDGSSF	GGGTGL

*Protein Analyses.* Structural models of A9UPE9 PDZ were created using the Robetta online server [64]. A structural homology model of A9UPE9 SH3 was created using SwissModel [94–96]. Sequence alignments were performed using BLAST and T-coffee [65,97]. All structure figures were created using PyMOL.

## 5. Conclusions

Here we provide an experimental pipeline to predict endogenous targets of a previously uncharacterized PDZ domain with no homology to any human protein. In our approach, we use structural conservation and sequence analyses to determine positional selectivity determinants, and successfully identify two choanoflagellate sequences with endogenously relevant binding affinities. We argue that this type of approach is applicable to other peptide-binding domains, e.g., SH2 and SH3 domains. These experiments have the potential to provide important insights into signaling pathways in choanoflagellates and other organisms, allowing us to better understand the origins of multicellularity.

**Author Contributions:** Conceptualization, H.A.W., J.M., F.J., and J.F.A.; methodology, F.J. and J.F.A.; software, J.M., B.V., and F.J.; validation, H.A.W. and J.M.; formal analysis, J.F.A.; investigation, H.A.W., J.M., B.V., K.A.E.A., F.L.; resources, F.J. and J.F.A.; data curation, J.F.A.; writing—original draft preparation, J.F.A.; writing—review and editing, H.A.W., J.M., F.J., and J.F.A.; visualization, H.A.W., J.M., F.J., and J.F.A.; supervision, F.J. and J.F.A.; project administration, J.F.A.; funding acquisition, J.F.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by NSF CHE-1904711 to JFA.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** All computational programs developed in this work are available in a GitHub repository, accessible from <https://github.com/joshmyersdean/damm>.

**Acknowledgments:** The authors would like to thank all members of the Amacher and Jagodzinski labs for helpful discussions and assistance. In addition, CSCI 474 Bioinformatics (Fall 2020) including Richard Carlstrom, Akimi Green, Kensington Hartman, D. Alex Johnson, Lauren Johnson, Sean Kearney, and Justin Svendsen provided important contributions to early stages of this project.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

**Sample Availability:** Samples of the constructs used in this study are available by the authors.

## References

1. Woods, D. F.; Bryant, P. J. Molecular cloning of the lethal(1)discs large-1 oncogene of *Drosophila*. *Dev. Biol.* **1989**, *134*, 222–235.
2. Bryant, P. J.; Watson, K. L.; Justice, R. W.; Woods, D. F. Tumor suppressor genes encoding proteins required for cell interactions and signal transduction in *Drosophila*. *Dev. Suppl.* **1993**, 239–249.
3. Woods, D. F.; Bryant, P. J. The discs-large tumor suppressor gene of *Drosophila* encodes a guanylate kinase homolog localized at septate junctions. *Cell* **1991**, *66*, 451–464.
4. Cho, K. O.; Hunt, C. A.; Kennedy, M. B. The rat brain postsynaptic density fraction contains a homolog of the *Drosophila* discs-large tumor suppressor protein. *Neuron* **1992**, *9*, 929–942.
5. Amacher, J. F.; Brooks, L.; Hampton, T. H.; Madden, D. R. Specificity in PDZ-peptide interaction networks: Computational analysis and review. *Journal of Structural Biology: X* **2020**, *4*, 100022.
6. Nourry, C.; Grant, S. G. N.; Borg, J.-P. PDZ domain proteins: plug and play! *Sci STKE* **2003**, *2003*, RE7.
7. Harris, B. Z.; Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* **2001**, *114*, 3219–3231.
8. Ponting, C. P. Evidence for PDZ domains in bacteria, yeast, and plants. *Protein Sci.* **1997**, *6*, 464–468.
9. Muley, V. Y.; Akhter, Y.; Galande, S. PDZ domains across the microbial world: molecular link to the proteases, stress response, and protein synthesis. *Genome Biol. Evol.* **2019**, *11*, 644–659.
10. Ren, A.; Zhang, W.; Yarlagadda, S.; Sinha, C.; Arora, K.; Moon, C.-S.; Naren, A. P. MAST205 competes with cystic fibrosis transmembrane conductance regulator (CFTR)-associated ligand for binding to CFTR to regulate CFTR-mediated fluid transport. *J. Biol. Chem.* **2013**, *288*, 12325–12334.

11. Dunn, H. A.; Ferguson, S. S. G. PDZ Protein Regulation of G Protein-Coupled Receptor Trafficking and Signaling Pathways. *Mol. Pharmacol.* **2015**, *88*, 624–639.
12. Cushing, P. R.; Fellows, A.; Villone, D.; Boisguérin, P.; Madden, D. R. The relative binding affinities of PDZ partners for CFTR: a biochemical basis for efficient endocytic recycling. *Biochemistry* **2008**, *47*, 10084–10098.
13. Raghuram, V.; Mak, D. O.; Foskett, J. K. Regulation of cystic fibrosis transmembrane conductance regulator single-channel gating by bivalent PDZ-domain-mediated interaction. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 1300–1305.
14. Cheng, J.; Moyer, B. D.; Milewski, M.; Loffing, J.; Ikeda, M.; Mickle, J. E.; Cutting, G. R.; Li, M.; Stanton, B. A.; Guggino, W. B. A Golgi-associated PDZ domain protein modulates cystic fibrosis transmembrane regulator plasma membrane expression. *J. Biol. Chem.* **2002**, *277*, 3520–3529.
15. Paasche, J. D.; Attramadal, T.; Kristiansen, K.; Oksvold, M. P.; Johansen, H. K.; Huitfeldt, H. S.; Dahl, S. G.; Attramadal, H. Subtype-specific sorting of the ETA endothelin receptor by a novel endocytic recycling signal for G protein-coupled receptors. *Mol. Pharmacol.* **2005**, *67*, 1581–1590.
16. Olson, P. A.; Tkatch, T.; Hernandez-Lopez, S.; Ulrich, S.; Ilijic, E.; Mugnaini, E.; Zhang, H.; Bezprozvanny, I.; Surmeier, D. J. G-protein-coupled receptor modulation of striatal CaV1.3 L-type Ca<sup>2+</sup> channels is dependent on a Shank-binding domain. *J. Neurosci.* **2005**, *25*, 1050–1062.
17. Rogan, M. P.; Stoltz, D. A.; Hornick, D. B. Cystic fibrosis transmembrane conductance regulator intracellular processing, trafficking, and opportunities for mutation-specific treatment. *Chest* **2011**, *139*, 1480–1490.
18. Karthikeyan, S.; Leung, T.; Ladias, J. A. Structural basis of the Na<sup>+</sup>/H<sup>+</sup> exchanger regulatory factor PDZ1 interaction with the carboxyl-terminal region of the cystic fibrosis transmembrane conductance regulator. *J. Biol. Chem.* **2001**, *276*, 19683–19686.
19. Guggino, W. B.; Stanton, B. A. New insights into cystic fibrosis: molecular switches that regulate CFTR. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 426–436.
20. Jeanneteau, F.; Diaz, J.; Sokoloff, P.; Griffon, N. Interactions of GIPC with dopamine D2, D3 but not D4 receptors define a novel mode of regulation of G protein-coupled receptors. *Mol. Biol. Cell* **2004**, *15*, 696–705.
21. Jeanneteau, F.; Guillin, O.; Diaz, J.; Griffon, N.; Sokoloff, P. GIPC recruits GAIP (RGS19) to attenuate dopamine D2 receptor signaling. *Mol. Biol. Cell* **2004**, *15*, 4926–4937.
22. Gujral, T. S.; Karp, E. S.; Chan, M.; Chang, B. H.; MacBeath, G. Family-wide investigation of PDZ domain-mediated protein-protein interactions implicates  $\beta$ -catenin in maintaining the integrity of tight junctions. *Chem. Biol.* **2013**, *20*, 816–827.
23. Cheng, B.; Montmasson, M.; Terradot, L.; Rousselle, P. Syndecans as Cell Surface Receptors in Cancer Biology. A Focus on their Interaction with PDZ Domain Proteins. *Front. Pharmacol.* **2016**, *7*, 10.
24. Ebnet, K.; Schulz, C. U.; Meyer Zu Brickwedde, M. K.; Pendl, G. G.; Vestweber, D. Junctional adhesion molecule interacts with the PDZ domain-containing proteins AF-6 and ZO-1. *J. Biol. Chem.* **2000**, *275*, 27979–27988.
25. Rademacher, N.; Schmerl, B.; Lardong, J. A.; Wahl, M. C.; Shoichet, S. A. MPP2 is a postsynaptic MAGUK scaffold protein that links SynCAM1 cell adhesion molecules to core components of the postsynaptic density. *Sci. Rep.* **2016**, *6*, 35283.
26. Cui, H.; Hayashi, A.; Sun, H.-S.; Belmares, M. P.; Cobey, C.; Phan, T.; Schweizer, J.; Salter, M. W.; Wang, Y. T.; Tasker, R. A.; Garman, D.; Rabinowitz, J.; Lu, P. S.; Tymianski, M. PDZ protein interactions underlying NMDA receptor-mediated excitotoxicity and neuroprotection by PSD-95 inhibitors. *J. Neurosci.*

2007, 27, 9901–9915.

27. Kim, E.; Sheng, M. PDZ domain proteins of synapses. *Nat. Rev. Neurosci.* **2004**, *5*, 771–781.
28. Christensen, N. R.; Čalyševa, J.; Fernandes, E. F. A.; Lüchow, S.; Clemmensen, L. S.; Haugaard-Kedström, L. M.; Strømgaard, K. PDZ domains as drug targets. *Adv. Therap.* **2019**, *2*, 1800143.
29. Thorsen, T. S.; Madsen, K. L.; Rebola, N.; Rathje, M.; Anggono, V.; Bach, A.; Moreira, I. S.; Stühr-Hansen, N.; Dyhring, T.; Peters, D.; Beuming, T.; Haganir, R.; Weinstein, H.; Mülle, C.; Strømgaard, K.; Rønn, L. C. B.; Gether, U. Identification of a small-molecule inhibitor of the PICK1 PDZ domain that inhibits hippocampal LTP and LTD. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 413–418.
30. Chen, X.; Longgood, J. C.; Michnoff, C.; Wei, S.; Frantz, D. E.; Bezprozvanny, L. High-throughput screen for small molecule inhibitors of Mint1-PDZ domains. *Assay Drug Dev Technol* **2007**, *5*, 769–783.
31. Kundu, R.; Cushing, P. R.; Popp, B. V.; Zhao, Y.; Madden, D. R.; Ball, Z. T. Hybrid organic-inorganic inhibitors of a PDZ interaction that regulates the endocytic fate of CFTR. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 7217–7220.
32. Cushing, P. R.; Vouilleme, L.; Pellegrini, M.; Boisguerin, P.; Madden, D. R. A stabilizing influence: CAL PDZ inhibition extends the half-life of  $\Delta F508$ -CFTR. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 9907–9911.
33. Javier, R. T.; Rice, A. P. Emerging theme: cellular PDZ proteins as common targets of pathogenic viruses. *J. Virol.* **2011**, *85*, 11544–11556.
34. James, C. D.; Roberts, S. Viral Interactions with PDZ Domain-Containing Proteins-An Oncogenic Trait? *Pathogens* **2016**, *5*.
35. Lee, C.; Laimins, L. A. Role of the PDZ domain-binding motif of the oncoprotein E6 in the pathogenesis of human papillomavirus type 31. *J. Virol.* **2004**, *78*, 12366–12377.
36. Mischo, A.; Ohlenschläger, O.; Hortschansky, P.; Ramachandran, R.; Görlach, M. Structural insights into a wildtype domain of the oncoprotein E6 and its interaction with a PDZ domain. *PLoS One* **2013**, *8*, e62584.
37. Pim, D.; Bergant, M.; Boon, S. S.; Ganti, K.; Kranjec, C.; Massimi, P.; Subbaiah, V. K.; Thomas, M.; Tomaić, V.; Banks, L. Human papillomaviruses and the specificity of PDZ domain targeting. *FEBS J.* **2012**, *279*, 3530–3537.
38. Ivarsson, Y.; Arnold, R.; McLaughlin, M.; Nim, S.; Joshi, R.; Ray, D.; Liu, B.; Teyra, J.; Pawson, T.; Moffat, J.; Li, S. S.-C.; Sidhu, S. S.; Kim, P. M. Large-scale interaction profiling of PDZ domains through proteomic peptide-phage display using human and viral phage peptidomes. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 2542–2547.
39. Shepley-McTaggart, A.; Sagum, C. A.; Oliva, I.; Rybakovsky, E.; DiGuilio, K.; Liang, J.; Bedford, M. T.; Cassel, J.; Sudol, M.; Mullin, J. M.; Harty, R. N. SARS-CoV-2 Envelope (E) protein interacts with PDZ-domain-2 of host tight junction protein ZO1. *PLoS One* **2021**, *16*, e0251955.
40. Javorsky, A.; Humbert, P. O.; Kvansakul, M. Structural basis of coronavirus E protein interactions with human PALS1 PDZ domain. *Commun. Biol.* **2021**, *4*, 724.
41. Toto, A.; Ma, S.; Malagrino, F.; Visconti, L.; Pagano, L.; Stromgaard, K.; Gianni, S. Comparing the binding properties of peptides mimicking the Envelope protein of SARS-CoV and SARS-CoV-2 to the PDZ domain of the tight junction-associated PALS1 protein. *Protein Sci.* **2020**, *29*, 2038–2042.
42. Caillet-Saguy, C.; Durbesson, F.; Rezelj, V. V.; Gogl, G.; Dinh Tran, Q.; Twizere, J.-C.; Vignuzzi, M.; Vincentelli, R.; Wolff, N. Host PDZ-containing proteins targeted by SARS-Cov-2. *FEBS J.* **2021**.
43. Zhang, Q.; Geftter, J.; Sneddon, W. B.; Mamonova, T.; Friedman, P. A. ACE2 interaction with cytoplasmic PDZ protein enhances SARS-CoV-2 invasion. *iScience* **2021**, *24*, 102770.

44. Songyang, Z.; Fanning, A. S.; Fu, C.; Xu, J.; Marfatia, S. M.; Chishti, A. H.; Crompton, A.; Chan, A. C.; Anderson, J. M.; Cantley, L. C. Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. *Science* **1997**, *275*, 73–77.
45. Amacher, J. F.; Cushing, P. R.; Brooks, L.; Boisguerin, P.; Madden, D. R. Stereochemical preferences modulate affinity and selectivity among five PDZ domains that bind CFTR: comparative structural and sequence analyses. *Structure* **2014**, *22*, 82–93.
46. Amacher, J. F.; Cushing, P. R.; Bahl, C. D.; Beck, T.; Madden, D. R. Stereochemical determinants of C-terminal specificity in PDZ peptide-binding domains: a novel contribution of the carboxylate-binding loop. *J. Biol. Chem.* **2013**, *288*, 5114–5126.
47. Tonikian, R.; Zhang, Y.; Sazinsky, S. L.; Currell, B.; Yeh, J.-H.; Reva, B.; Held, H. A.; Appleton, B. A.; Evangelista, M.; Wu, Y.; Xin, X.; Chan, A. C.; Seshagiri, S.; Lasky, L. A.; Sander, C.; Boone, C.; Bader, G. D.; Sidhu, S. S. A specificity map for the PDZ domain family. *PLoS Biol.* **2008**, *6*, e239.
48. Gao, M.; Mackley, I. G. P.; Mesbahi-Vasey, S.; Bamonte, H. A.; Struyvenberg, S. A.; Landolt, L.; Pederson, N. J.; Williams, L. I.; Bahl, C. D.; Brooks, L.; Amacher, J. F. Structural characterization and computational analysis of PDZ domains in *Monosiga brevicollis*. *Protein Sci.* **2020**, *29*, 2226–2244.
49. King, N.; Westbrook, M. J.; Young, S. L.; Kuo, A.; Abedin, M.; Chapman, J.; Fairclough, S.; Hellsten, U.; Isogai, Y.; Letunic, I.; Marr, M.; Pincus, D.; Putnam, N.; Rokas, A.; Wright, K. J.; Zuzow, R.; Dirks, W.; Good, M.; Goodstein, D.; Lemons, D.; Li, W.; Lyons, J. B.; Morris, A.; Nichols, S.; Richter, D. J.; Salamov, A.; Sequencing, J. G. I.; Bork, P.; Lim, W. A.; Manning, G.; Miller, W. T.; McGinnis, W.; Shapiro, H.; Tjian, R.; Grigoriev, I. V.; Rokhsar, D. The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* **2008**, *451*, 783–788.
50. Ruiz-Trillo, I.; Roger, A. J.; Burger, G.; Gray, M. W.; Lang, B. F. A phylogenomic investigation into the origin of metazoa. *Mol. Biol. Evol.* **2008**, *25*, 664–672.
51. Laundon, D.; Larson, B. T.; McDonald, K.; King, N.; Burkhardt, P. The architecture of cell differentiation in choanoflagellates and sponge choanocytes. *PLoS Biol.* **2019**, *17*, e3000226.
52. Levin, T. C.; Greaney, A. J.; Wetzel, L.; King, N. The Rosetteless gene controls development in the choanoflagellate *S. rosetta*. *Elife* **2014**, *3*.
53. Dayel, M. J.; Alegado, R. A.; Fairclough, S. R.; Levin, T. C.; Nichols, S. A.; McDonald, K.; King, N. Cell differentiation and morphogenesis in the colony-forming choanoflagellate *Salpingoeca rosetta*. *Dev. Biol.* **2011**, *357*, 73–82.
54. Valgardson, J.; Cosbey, R.; Houser, P.; Rupp, M.; Van Bronkhorst, R.; Lee, M.; Jagodzinski, F.; Amacher, J. F. MotifAnalyzer-PDZ: A computational program to investigate the evolution of PDZ-binding target specificity. *Protein Sci.* **2019**, *28*, 2127–2143.
55. Hui, S.; Xing, X.; Bader, G. D. Predicting PDZ domain mediated protein interactions from structure. *BMC Bioinformatics* **2013**, *14*, 27.
56. Murciano-Calles, J.; McLaughlin, M. E.; Erijman, A.; Hooda, Y.; Chakravorty, N.; Martinez, J. C.; Shifman, J. M.; Sidhu, S. S. Alteration of the C-terminal ligand specificity of the erbin PDZ domain by allosteric mutational effects. *J. Mol. Biol.* **2014**, *426*, 3500–3508.
57. Elkins, J. M.; Papagrigoriou, E.; Berridge, G.; Yang, X.; Phillips, C.; Gileadi, C.; Savitsky, P.; Doyle, D. A. Structure of PICK1 and other PDZ domains obtained with the help of self-binding C-terminal extensions. *Protein Sci.* **2007**, *16*, 683–694.
58. Skelton, N. J.; Koehler, M. F. T.; Zobel, K.; Wong, W. L.; Yeh, S.; Pisabarro, M. T.; Yin, J. P.; Lasky, L. A.; Sidhu, S. S. Origins of PDZ domain ligand specificity. Structure determination and mutagenesis of the Erbin PDZ domain. *J. Biol. Chem.* **2003**, *278*, 7645–7654.



59. Tonikian, R.; Zhang, Y.; Boone, C.; Sidhu, S. S. Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. *Nat. Protoc.* **2007**, *2*, 1368–1386.
60. Vouilleme, L.; Cushing, P. R.; Volkmer, R.; Madden, D. R.; Boisguerin, P. Engineering peptide inhibitors to overcome PDZ binding promiscuity. *Angew. Chem. Int. Ed. Engl.* **2010**, *49*, 9912–9916.
61. Liu, Y.; Henry, G. D.; Hegde, R. S.; Baleja, J. D. Solution structure of the hDlg/SAP97 PDZ2 domain and its mechanism of interaction with HPV-18 papillomavirus E6 protein. *Biochemistry* **2007**, *46*, 10864–10874.
62. Rodzli, N. A.; Lockhart-Cairns, M. P.; Levy, C. W.; Chipperfield, J.; Bird, L.; Baldock, C.; Prince, S. M. The Dual PDZ Domain from Postsynaptic Density Protein 95 Forms a Scaffold with Peptide Ligand. *Biophys. J.* **2020**, *119*, 667–689.
63. Nomme, J.; Antanasijevic, A.; Caffrey, M.; Van Itallie, C. M.; Anderson, J. M.; Fanning, A. S.; Lavie, A. Structural Basis of a Key Factor Regulating the Affinity between the Zonula Occludens First PDZ Domain and Claudins. *J. Biol. Chem.* **2015**, *290*, 16595–16606.
64. Kim, D. E.; Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004**, *32*, W526–31.
65. Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
66. Ganti, K.; Broniarczyk, J.; Manoubi, W.; Massimi, P.; Mittal, S.; Pim, D.; Szalmas, A.; Thatte, J.; Thomas, M.; Tomaić, V.; Banks, L. The human papillomavirus E6 PDZ binding motif: from life cycle to malignancy. *Viruses* **2015**, *7*, 3530–3551.
67. Thomas, M.; Myers, M. P.; Massimi, P.; Guarnaccia, C.; Banks, L. Analysis of Multiple HPV E6 PDZ Interactions Defines Type-Specific PDZ Fingerprints That Predict Oncogenic Potential. *PLoS Pathog.* **2016**, *12*, e1005766.
68. Webb Strickland, S.; Brimer, N.; Lyons, C.; Vande Pol, S. B. Human Papillomavirus E6 interaction with cellular PDZ domain proteins modulates YAP nuclear localization. *Virology* **2018**, *516*, 127–138.
69. Zhang, Z.; Li, H.; Chen, L.; Lu, X.; Zhang, J.; Xu, P.; Lin, K.; Wu, G. Molecular basis for the recognition of adenomatous polyposis coli by the Discs Large 1 protein. *PLoS One* **2011**, *6*, e23507.
70. Adam, M. G.; Berger, C.; Feldner, A.; Yang, W.-J.; Wüsthube-Lausch, J.; Herberich, S. E.; Pinder, M.; Gesierich, S.; Hammes, H.-P.; Augustin, H. G.; Fischer, A. Synaptojanin-2 binding protein stabilizes the Notch ligands DLL1 and DLL4 and inhibits sprouting angiogenesis. *Circ. Res.* **2013**, *113*, 1206–1218.
71. Hartmann, C.; Schwietzer, Y. A.; Kummer, D.; Kirschnick, N.; Hoppe, E.; Thüning, E.-M.; Glaesner-Ebnet, M.; Brinkmann, F.; Gerke, V.; Reuter, S.; Nakayama, M.; Ebnet, K. The mitochondrial outer membrane protein SYNJ2BP interacts with the cell adhesion molecule TMIGD1 and can recruit it to mitochondria. *BMC Mol and Cell Biol* **2020**, *21*, 30.
72. Carr, H. S.; Chang, J. T.; Frost, J. A. The PDZ Domain Protein SYNJ2BP Regulates GRK-Dependent Sst2A Phosphorylation and Downstream MAPK Signaling. *Endocrinology* **2021**, *162*.
73. Sadowski, I.; Stone, J. C.; Pawson, T. A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol.* **1986**, *6*, 4396–4408.
74. Sadowski, I.; Pawson, T. Catalytic and non-catalytic domains of the Fujinami sarcoma virus P130gag-fps protein-tyrosine kinase distinguished by the expression of v-fps polypeptides in *Escherichia coli*. *Oncogene* **1987**, *1*, 181–191.
75. Stahl, M. L.; Ferez, C. R.; Kelleher, K. L.; Kriz, R. W.; Knopf, J. L. Sequence similarity of phospholipase C with the non-catalytic region of src. *Nature* **1988**, *332*, 269–272.

76. Liu, B. A.; Nash, P. D. Evolution of SH2 domains and phosphotyrosine signalling networks. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **2012**, *367*, 2556–2573.
77. Teyra, J.; Huang, H.; Jain, S.; Guan, X.; Dong, A.; Liu, Y.; Tempel, W.; Min, J.; Tong, Y.; Kim, P. M.; Bader, G. D.; Sidhu, S. S. Comprehensive Analysis of the Human SH3 Domain Family Reveals a Wide Variety of Non-canonical Specificities. *Structure* **2017**, *25*, 1598–1610.e3.
78. Waksman, G.; Kominos, D.; Robertson, S. C.; Pant, N.; Baltimore, D.; Birge, R. B.; Cowburn, D.; Hanafusa, H.; Mayer, B. J.; Overduin, M.; Resh, M. D.; Rios, C. B.; Silverman, L.; Kuriyan, J. Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **1992**, *358*, 646–653.
79. Liu, B. A.; Engelmann, B. W.; Nash, P. D. The language of SH2 domain interactions defines phosphotyrosine-mediated signal transduction. *FEBS Lett.* **2012**, *586*, 2597–2605.
80. Liu, B. A.; Jablonowski, K.; Shah, E. E.; Engelmann, B. W.; Jones, R. B.; Nash, P. D. SH2 domains recognize contextual peptide sequence information to determine selectivity. *Mol. Cell Proteomics* **2010**, *9*, 2391–2404.
81. Waksman, G.; Shoelson, S. E.; Pant, N.; Cowburn, D.; Kuriyan, J. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **1993**, *72*, 779–790.
82. Feng, S.; Kasahara, C.; Rickles, R. J.; Schreiber, S. L. Specific interactions outside the proline-rich core of two classes of Src homology 3 ligands. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 12408–12415.
83. Schlessinger, J. SH2/SH3 signaling proteins. *Curr. Opin. Genet. Dev.* **1994**, *4*, 25–30.
84. Lemmon, M. A.; Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **2010**, *141*, 1117–1134.
85. Pawson, T. Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* **2004**, *116*, 191–203.
86. Pawson, T.; Gish, G. D.; Nash, P. SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol.* **2001**, *11*, 504–511.
87. Pawson, T.; Olivier, P.; Rozakis-Adcock, M.; McGlade, J.; Henkemeyer, M. Proteins with SH2 and SH3 domains couple receptor tyrosine kinases to intracellular signalling pathways. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **1993**, *340*, 279–285.
88. King, N.; Carroll, S. B. A receptor tyrosine kinase from choanoflagellates: molecular insights into early animal evolution. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 15032–15037.
89. Li, W.; Scarlata, S.; Miller, W. T. Evidence for convergent evolution in the signaling properties of a choanoflagellate tyrosine kinase. *Biochemistry* **2009**, *48*, 5180–5186.
90. Manning, G.; Young, S. L.; Miller, W. T.; Zhai, Y. The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9674–9679.
91. Miller, W. T. Tyrosine kinase signaling and the emergence of multicellularity. *Biochim. Biophys. Acta* **2012**, *1823*, 1053–1057.
92. Li, W.; Young, S. L.; King, N.; Miller, W. T. Signaling properties of a non-metazoan Src kinase and the evolutionary history of Src negative regulation. *J. Biol. Chem.* **2008**, *283*, 15491–15501.
93. Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423.
94. Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL workspace: a web-based

- environment for protein structure homology modelling. *Bioinformatics* **2006**, *22*, 195–201.
95. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303.
96. Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **2009**, *4*, 1–13.
97. Notredame, C.; Higgins, D. G.; Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **2000**, *302*, 205–217.