The protist, *Monosiga brevicollis*, has a tyrosine kinase signaling network more elaborate and diverse than found in any known metazoan

Gerard Manning*†, Susan L. Young‡, W. Todd Miller§, and Yufeng Zhai*

*Razavi Newman Center for Bioinformatics, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037; ‡Department of Molecular and Cell Biology and Center for Integrative Genomics, University of California, Berkeley, CA 94720; and §Department of Physiology and Biophysics, Stony Brook University, Stony Brook, NY 11794

Edited by Tony Hunter, Salk Institute for Biological Studies, La Jolla, CA, and approved April 28, 2008 (received for review February 11, 2008)

Tyrosine kinase signaling has long been considered a hallmark of intercellular communication, unique to multicellular animals. Our genomic analysis of the unicellular choanoflagellate *Monosiga brevicollis* discovers a remarkable count of 128 tyrosine kinases, 38 tyrosine phosphatases, and 123 phosphotyrosine (pTyr)-binding SH2 proteins, all higher counts than seen in any metazoan. This elaborate signaling network shows little orthology to metazoan counterparts yet displays many innovations reminiscent of metazoans. These include extracellular domains structurally related to those of metazoan receptor kinases, alternative methods for membrane anchoring and phosphotyrosine interaction in cytoplasmic kinases, and domain combinations that link kinases to small GTPase signaling and transcription. These proteins also display a wealth of combinations of known signaling domains. This uniquely divergent and elaborate signaling network illuminates the early evolution of pTyr signaling, explores innovative ways to traverse the cellular signaling circuitry, and shows extensive convergent evolution, highlighting pervasive constraints on pTyr signaling.

**Results**

Our analysis of the draft *Monosiga* genome predicts 128 TKs within a total kinome of ~380 protein kinases (http://kinase.com/kinbase). Extensive gene model curation and selected cDNA and genome resequencing allowed us to improve predictions for 102 of these sequences, although several fragments and likely imperfect predictions remain. These constitute the largest known tyrosine kinome and make up over twice the fraction of the proteome than that of any metazoan (6–9), a startling result for a unicellular organism. Sequence analysis of the kinase domain and other regions clusters these TKs into 22 families and 26 singletons (Fig. 1). Their scope is paralleled by their diversity: when compared with metazoan TKs by pairwise and multiple sequence alignment and family profile–profile alignments, the only clearly identifiable specific homologs were of the Src subgroup kinases (Src, Csk, Abl, and Tec).

Receptor TKs (RTKs). Eighty-eight RTKs are predicted, based on predicted signal peptides and transmembrane (TM) regions, known extracellular domains, and paralogy. Most are typical type I TM proteins, but two are multispan (six to nine adjacent predicted TMs), including one encoding transporter domain [supporting information (SI) Fig. S1]. Seventy-three RTKs belong to 15 families, none of which have obvious metazoan orthologs, although kinase domain profile–profile alignments do show weakly specific similarity between RTKB and RTKC families and the metazoan FGFR and Eph families, respectively. Their domain organization is often similar to that of metazoans, whether due to common origin or convergent evolution (Table S1, Fig. 2). For instance, *Monosiga* lacks the Ig domains found in many metazoan RTKs, but 15 *Monosiga* RTKs have divergent repeats similar to hyalinit (HYR) domains, which in turn are predicted to be structurally related to Ig and FN3 domains (10). Similarly, 21 *Monosiga* RTKs have cysteine-rich extracellular repeats and several families of CxC motifs. These are weakly similar to the TNFR and furin-like domains of some metazoan RTKs. Variant EGF-like domains are also seen (Table S1). Several of these domains are found in other predicted receptor and secreted *Monosiga* proteins. For instance, the *Monosiga*-specific RM1 motif is repeated 8–13 times in three RTKs (SI Text) and in 40 other proteins, most of which are predicted to be secreted.

Cytoplasmic TKs (CTKs). Most CTKs are associated with membrane and pTyr binding, and, as in metazoans, are likely to transduce signals from activated receptors, although frequently with distinct domain combinations. Twenty-nine of the 40 CTKs fall into eight families, seven of which also contain SH2 or phosphotyrosine binding (PTB) domains (Fig. 2, Fig. S2). These include homologs of all four Src subgroup families, based on presence of SH2 and SH3 domains and on kinase domain sequence similarity, which averages


The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the kinase. com database, http://kinase.com/kinbase/FastaFiles (accession nos. Mbre0001–Mbre0128).

See Commentary on page 9453.

†To whom correspondence should be addressed. E-mail: manning@salk.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/ 0801314105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA
60% identity to their closest metazoan homologs, compared with <50% for any other Monosiga kinase. As in metazoans, all but Csk have an activation loop phosphorylation site, and all four Src family kinases have conserved Csk phosphorylation sites at their C termini. In Monosiga ovata, Csk has been shown to phosphorylate and partially repress Src activity through this site (5).

Three of the four Srcs have predicted membrane-anchoring myristoylation sites, indicating that they function proximal to receptors, as with their metazoan counterparts. Curiously, the fourth replaces this with a predicted lipid-binding C2 domain that suggests a novel mechanism of membrane targeting, perhaps similar to the PH domain of Tec kinases (Fig. S2).

Other CTK families also have pTyr-binding domains and may be downstream of RTKs. The two FYTK kinases have SH2 and inositol lipid-binding FYVE domains, one CTKA kinase has SH2 and PH domains, and 10 of the 15 HMTK (HM-motif TK) kinases have PTB domains (Fig. S1). Although FYVE and PTB domains have not previously been seen in TKs, they may function analogously to the membrane targeting (PH, myristoylation) and pTyr-binding (SH2) domains of Src subgroup kinases.

Several RTKs contain predicted Src phosphorylation and SH2-binding sites, most notably at four conserved tyrosines in the RM2 motif within the tail of several RTKB kinases (Fig. S1, SI Text). We tested biochemical activity of Monosiga Src1 on peptides generated from two copies of this motif from RTKB2, along with Monosiga STAT (a predicted Src substrate) and an optimal vertebrate Src substrate. All showed distinct activity, but the specific activity toward the RTKB2 peptides under these conditions was 6-fold higher (Fig. 3). Kinetic analysis of phosphorylation showed that RTKB2–1 had a $k_{cat}$ of 97.4 min$^{-1}$ and a $K_m$ of 280 µM, whereas the c-Src optimal peptide gave $k_{cat}$ = 6.5 min$^{-1}$ and $K_m$ = 90 µM. Thus, specific recognition of RTKB2–1 by Src1 is driven primarily by a high maximal velocity toward this substrate. These data raise the possibility that the RTKB tail is a Src substrate, thus linking RTK and CTK signaling, as in metazoans, and that initial auto-phosphorylation of one of these sites by the RTK recruits Src for further phosphorylation.

Kinase Domain Conservation and Catalytic Activity. Given their ancient divergence, we tested whether Monosiga TK domains had unique sequence features. Comparison of all Monosiga TK domains to all human, Drosophila, and Caenorhabditis elegans TK domains by HMM profiles shows a remarkable similarity (Fig. S3), with no clear difference in the conservation profile at any part of the domain. This suggests that TKs in both lineages are under similar constraints, and that their common ancestor had already taken on a “mature” TK structure. Most appear to be activated by phosphorylation, because 103 TKs conserve one or more tyrosines in the phosphorylatable region of the activation loop (Dataset S1).

In other species, several RTKs have lost key catalytic residues and are thought to act as scaffolds or coreceptors, including the EGF receptor ErbB3 and several Eph receptors (7). By this measure, 13 Monosiga TKs are inactive (Dataset S1). Most belong to the RTKB or RTKM families or are unclassified. Unlike in human, three of the inactive Monosiga kinases appear to be cytoplasmic.

Nine kinases have dual catalytic domains, including the six RTKE receptors and the two CTKA cytoplasmic kinases. In all cases, one of the two domains is predicted to be catalytically inactive and is usually very divergent or fragmentary. This situation is analogous to but distinct from metazoan Jak kinases, whose inactive second kinase domains are autoinhibitory (11).
Other Phosphotyrosine Signaling Proteins. The richness and diversity of tyrosine kinases are reflected in downstream pTyr-dependent proteins. Conventional tyrosine-specific phosphatases (PTP) and pTyr-binding domains (SH2, PTB) are also greatly expanded in number and domain complexity when compared with yeast, Dictyostelium, or Tetrahymena and surpass even the human counts for PTP and SH2 proteins (Table 1). As with TKs, we see limited orthology to metazoans, tremendous diversity and several recurrent themes and variations in domain architecture (Fig. 4).

Unlike the few other unicellular PTPs, 4 of the 39 Monosiga PTPs have clear human orthologs. These include SHP, PTPN13 (PTP-BAS), PTP23 (HD-PTP), and PTP N3/N4. Curiously, Drosophila lacks PTPN13, and both Drosophila and C. elegans lack PTPN23, so, although ancient, these are not evolutionarily indispensable. Both SHP and PTPN13 have been shown to dephosphorylate Src in mammals (12). As in metazoans, some PTPs appear to be catalytically inactive, and four have lost their HCxxxxxR active site motifs (Fig. S2).

By contrast, over one-fifth (26 of 123) of the SH2 proteins have human orthologs covering 15 classes (13) and all 11 major functional categories (Table 2, Fig. 4, Fig. S2). This indicates that much of the cellular pTyr-modulated circuitry was present in the unicellular common ancestor, despite the limited orthology in TKs and PTPs. These shared SH2 proteins mediate pTyr modulation of major signaling pathways, including Ras, Rho, Rac, and Cdc42 small GTPases, phospholipid and calcium signaling, transcription, cytoskeletal interactions, Src subgroup tyrosine kinase and SHP phosphatase signaling, and several adaptors and scaffolds. The remaining 98 SH2 proteins and 35 PTPs lack metazoan orthologs, but many have domain combinations that suggest common themes and the development of new circuits within a constrained set of

Table 1. Number of proteins with pTyr associated signaling domains in selected genomes

<table>
<thead>
<tr>
<th>Species</th>
<th>TK</th>
<th>PTP</th>
<th>SH2</th>
<th>PTB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetrahymena thermophila</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Dictyostelium discoideum</td>
<td>0</td>
<td>3</td>
<td>13 (14)</td>
<td>0</td>
</tr>
<tr>
<td>S. cerevisiae</td>
<td>0</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>M. brevicollis</td>
<td>128 (136)</td>
<td>39 (40)</td>
<td>123 (143)</td>
<td>20 (31)</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>33 (34)</td>
<td>16 (23)</td>
<td>28 (34)</td>
<td>10</td>
</tr>
<tr>
<td>Human</td>
<td>90 (94)</td>
<td>38 (50)</td>
<td>110 (120)</td>
<td>46 (51)</td>
</tr>
<tr>
<td>Human–Monosiga orthologs</td>
<td>4</td>
<td>4–5</td>
<td>19</td>
<td>1</td>
</tr>
</tbody>
</table>

Parentheses indicate domain count due to multidomain proteins. Human counts from RefSeq analysis and published studies (13, 16, 29).
signaling interactions. These include 10 receptor PTPs (rPTP), previously unique to metazoans, and 15 cases of a previously undescribed receptor-SH2 (rSH2) combination (Fig. S2). These bring to 103 the count of pTyr-linked receptors. Two rPTPs and three rSH2s are cadherins (2 PTP, 3 SH2), a class best known as metazoan cell adhesion proteins (4, 14). Other extracellular domains include the Monosiga-specific variant HYR and cysteine-rich regions also seen in RTKs and several more conventional extracellular domains (FN3, TIG, VWA, TSP, EGF). The single dual-domain rPTP is a possible homolog of the metazoan LAR family, but as with RTKs, the other rPTPs have no clear orthologs.

Several PTP and SH2 domains are fused to Class III myosins. This class was previously found only in combination with the NinaC subfamily of Ser/Thr kinases, which function in both phototransduction and hearing (15). Two PTPs are fused to the kinase–myosin combination, whereas seven SH2 domains are fused to the myosin but lack the kinase (Fig. 4, Fig. S2).

Many more PTP and SH2 proteins are linked to other signaling domains but in unique architectures or with no specific homology to human counterparts (Fig. S2). Partner domains consist mostly of protein, lipid, and calcium-binding adaptor modules, including SH2, SH3, PDZ, SAM, WW, C1, PH, ankyrin, and EF hand domains. Monosiga lacks orthologs of the metazoan SH2-RasGEF and SH2-C1-RhoGAP proteins but has unique and possibly analogous...
Table 2. Human orthologs of Monosiga SH2-containing proteins and their functions

<table>
<thead>
<tr>
<th>Ortholog</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crk</td>
<td>SH2-SH3 adaptor (RAP/RAC GEFs for adhesion)</td>
</tr>
<tr>
<td>Grb2</td>
<td>SH2-SH3 adaptor (SOS, Gab1-MAPK/PI3K)</td>
</tr>
<tr>
<td>SHP</td>
<td>PTP phosphatase: Src activator, RTK signaling</td>
</tr>
<tr>
<td>STAT</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>Cbl</td>
<td>Ubiquitination, receptor trafficking</td>
</tr>
<tr>
<td>PIK3R (p85)</td>
<td>Phosphoinositide signaling; PI3K regulatory subunit</td>
</tr>
<tr>
<td>PLCγ</td>
<td>Phospholipase: PI3K/Ca signaling adaptor</td>
</tr>
<tr>
<td>SHIP2</td>
<td>Phosphoinositide phosphatase</td>
</tr>
<tr>
<td>RASA1</td>
<td>Small GTPase: Ras adaptor</td>
</tr>
<tr>
<td>Rin</td>
<td>Small GTPase: CDC42 adaptor?</td>
</tr>
<tr>
<td>Vav</td>
<td>Small GTPase: Rho adaptor</td>
</tr>
<tr>
<td>Src/Abu/Csk/Tec</td>
<td>Src kinase signaling</td>
</tr>
<tr>
<td>TNS1</td>
<td>Cytoskeleton</td>
</tr>
<tr>
<td>SH2D4</td>
<td>Unknown</td>
</tr>
<tr>
<td>Sup26 h</td>
<td>Regulator of chromatin structure. Conserved in yeast, probably non-pTyr-binding</td>
</tr>
</tbody>
</table>

RasGEF-SH2-SH2 and RhoGAP-SH2 combinations. Similarly, Monosiga and metazoans have several nonorthologous proteins containing SH2 and PH domains, which may share related functions, although lacking obvious common ancestry. A few metazoan proteins have dual SH2 domains, but Monosiga has 27 multi-SH2 proteins, with up to six domains seen in a single protein. Many more domains (13 PTP, 28 SH2) consist of only one recognizable domain and probably include many fragmentary gene predictions.

PTB domains are prevalent in Monosiga and metazoans (Table 1) but are absent from lower organisms. Metazoan PTBs can bind peptides, phosphopeptides, or phospholipids (16). The specific ligands in Monosiga could not be predicted by sequence analysis, although domain combinations indicate several are associated with pTyr signaling. The PTB–kinase association in the HMTK family is novel, although reminiscent of the pairing of SH2 with the Src subgroup kinases and a likely case of convergent evolution. Monosiga has one PTB-SH2 protein that might be a homolog of the Shc adaptor, but other PTB proteins are unique and have no other domains.

Discussion

The Monosiga genome has revealed a treasure trove of diverse tyrosine kinases and associated pTyr signaling proteins. These demonstrate an unprecedented diversity relative to all known (metazoan) TK-based signaling yet reveal several common themes that suggest convergent evolution and a limited set of recurring molecular themes favored by signaling pathways. These data also highlight the unresolved puzzle of why a unicellular organism has such an elaborate signaling system based on external cues. Some choanoflagellates such as Proterospongia sp. do form colonies, and it may be that such a colonial ancestor drove the evolution of this system, yet it is clear from sequence conservation that pTyr signaling proteins continue to be essential for the current unicellular lifestyle of M. brevicollis. Possible functions include response to prey, predators, mates, and the abiotic environment.

Only the four Src-subgroup kinase families have detectable metazoan orthologs, although possible divergent homologs of FGFR and Eph RTKs may also be present. By contrast, most metazoan TK families are clearly visible in sponges (EGFR, FGFR, Eph, InsR, Ret, Musk, Sev, DDR RTKs and Jak, Syk, and Fer CTKs) (17, 18). This suggests that the choanoflagellate–metazoan common ancestor had a mature Src signaling and some RTKs, but that most metazoan TK families were established closer to the base of metazoans. The story is similar with PTPs, but the common ancestor apparently had a very extensive set of SH2 proteins, with many more classes invented within the choanoflagellate lineage. Extracellular domains evolve rapidly and can swap between families, such as in the RM1 domains found in both RTKα and RTKG1 kinases. Even more remarkably, one, or possibly two, of the receptor SH2 proteins have extracellular domains that are highly similar (>90% sequence identity) to RTKB3 and RTKB5, indicating these are recent fusions and suggesting a kinase–SH2 interaction by receptor heterodimerization. Monosiga-specific extracellular motifs are also seen in many other receptor proteins, including rPTPs and rSH2s, and secreted proteins, suggesting they have common ligands or may interact homotypically.

Intracellular domains are more evolutionarily stable and are dominated by pTyr, lipid, and protein interaction domains, as seen in metazoans. In addition, Monosiga pTyr-associated proteins have a strong association with the cytoskeleton, as evidenced by an unusual abundance of myosin, CAP, Gly, and calponin homology domains, and with a variety of small GTPase GAPS and GEFS.

Common themes and possible convergent evolution are seen in the domain structures of many pTyr signaling proteins. These include the swapping of a myristoylation site for a putative lipid-binding C2 domain in Src4, the common occurrence of HYR domains reminiscent of Ig domains, the use of cysteine-rich motifs in extracellular regions, and the use of PTB domains as membrane or phosphopeptide anchors that may be analogous to SH2 and myristoylation domains in Src kinases. The development of dual-domain and catalytically inactive kinases are also probably independent innovations in both lineages. In other cases, Monosiga proteins are associated with signaling domains not found in metazoans or found in a different architecture, indicating it has successfully explored new paths within signaling space.

Many of the combinatorial aspects of metazoan pTyr signaling are also found in Monosiga, including the widespread occurrence of activation loop phosphorylation sites, the likely phosphorylation of RTKs by Src and of Srcs by Csk, the predicted membrane localization of most CTKs, the conservation of most major classes of SH2 domain proteins, and the occurrence of many multi-SH2 proteins that may link distinct pTyr signals. Conversely, the absence of many metazoan components may allow experimental investigation of pathway alternatives, such as the likely specific activation of STAT by Src kinases in the absence of JAKs or the possible link between RTKs and MAPK signaling given the absence of any Raf kinase in Monosiga (1).

Future Prospects. This analysis of the draft Monosiga genome is surely just an exploratory step in understanding this elaborate and divergent network. The sequence divergence in Monosiga and the presence of many short exons hamper gene prediction. We manually improved 102 the kinase sequences over the genome predictions, but several are still clearly incomplete. Impending large-scale choanoflagellate EST and genome sequencing, including those for Proterospongia sp. and Monosiga ovata, will greatly improve our predictions and provide an evolutionary context. New proteomic technologies to identify TK substrate sites and signaling protein interactions (19–21) could quickly fill in much of the signaling network and allow large-scale comparisons to other systems. A greater understanding of pTyr signaling in choanoflagellates promises to reveal both variations on an important biological theme and commonalities that indicate common origin or convergent evolution.

Materials and Methods

Gene Identification. Protein sequences were predicted from release 1.0 of the Monosiga genome (1). Protein kinases, PTP, SH2, and PTB-containing proteins were identified by profile HMM searches against genomic, EST, and predicted gene sequences, using HMMer (http://hmmer.janelia.org), GeneWise (www.ebi.ac.uk/Wise2), and Gene Detective (a hardware-accelerated implementation of...
Cysteine-rich regions were identified by multiple overlapping hits to the Pfam manually. Overlapping domains from different profile families were merged by manual inspection and the MEME motif-finder (23), followed by HMM text, Figs. S3 and S4) with Global and Glocal models were performed with a using PRC (http://supfam.org/PRC).

Domain Profiles. HMM searches on Pfam, SMART, TIGR, and in-house HMM (SI Text, Figs. S3 and S4) with Global and Glocal models were performed with a hardware-accelerated DeCypher system (Active Motif). E value cutoffs of e < 10 were used to pick up repeated elements whose individual scores were very low. Sequence-level scores of e > 0.001 were discarded and scores of e > 1e-8 inspected manually. Overlapping domains from different profile families were merged. Cystein-rich regions were identified by multiple overlapping hits to the Pfam and SMART profiles GCC2.GCCS, TNFR.3, TNFR.c6, NCD3G, and to internal models for RMS, RMS, RMS, RT15, and RT15. Adjacent cysteine-rich regions were merged when separated by < 10 residues. Custom HMM profiles were built for several unique conserved regions, found by manual inspection and the MEME motif-finder (23), followed by HMM searches of Monosiga and GenBank protein and EST sequences to diversify the motifs found and occasional merging of adjacent motifs into gapped profiles. The Vav PH domain and Sup6 CSZ domain were detected by alignment to proteins with these domains but did not score significantly on the HMMs.

Signal peptides and TM segments were predicted by SignalP (24) and TM-HMM (25). TM that overlapped kinase domains or signal peptides were eliminated. Likely receptors that lacked either signal were subjected to gene repre-
diction and evaluated in part on the basis of these motifs; this may have lead to some overprediction of such motifs. Myristoylation sites were predicted with NMT (http://mendel.imp.ac.at/mrystate) (26).

Kinase Domain Conservation. The alignment of Monosiga and metazoan TK domain HMMs was built from a hand-edited alignment of all Monosiga TK kinase domains (Dataset S2) and an alignment of published TK domains of C. elegans, Drosophila, and human (6, 7, 9). The logo was generated with Logomat-M (27). Other Genomic Searches. Sequence files used for profile searches included Dictyostelium: “dicty.predicted.proteins” (http://dictybase.org, June 2007 download) Schizosaccharomyces cerevisiae: “SGD1.01.45.Known.pep” (wwwensembl.org); Drosophila BDGP 4.3.46 “all.pep” (wwwensembl.org); Tetrahymena thermophila “gene.prediction” (http://criate.org, May 2005 download), and human ReSeq proteins from GenBank, June 2007 download.

Sequenceing. Resequencing used either a M. brevicolis cDNA library (3) or cDNA. To generate cDNA, M. brevicolis (American Type Culture Collection 30154) was cultured with Enterobacter aerogenes in 25°C in natural seawater infused with cereal grass (S gilter) in 150–25-mm polystyrene dishes (Falco). Total RNA was extracted and DNase treated with RNasey Midi-prep kit (Qiagen). This was reverse-transcribed with an oligo(dt) primer (Invitrogen) and amplified using gene specific primers. PCR amplicons were cloned into pCR-Blunt II-TOPO (Invitrogen) and sequenced by using the vector specific primers M13F (5'- TTGAAAACGACGGCCAGT-3') and M13R (5’-AACACGATACCATG-3') or the gene-specific PCR primers.

Sct Phosphorylation Assay. Src1 was expressed and purified from insect cells (30). Phosphorylation assays were carried out in total volumes of 25 μl at 30°C, containing 50 mM Tris-HCl (pH 7.4), 10 mM MgCl2, 1 mg/ml BSA, 400 μM ATP (15 cpm/μmol), and 750 μM peptide sequence were: c-Src optimal, AEEEYEEFKAKK; MBStat, KKKASYGNYADIA; RTK82–2, AEEYEAJADK. Reactions were initiated by addition of kinase. Samples were centrifuged for 1 min, and 35-μl aliquots of the supernatants were spotted onto 2.1-cm phosphocellulose paper circles (27). The circles were washed three times with cold 0.5% phosphoric acid and once with acetone, dried, and counted dry in a liquid scintillation counter to measure incorporation of 32P into protein. Reactions were carried out in duplicate and are presented ± standard deviation. For kinetic measurements, reactions were carried out with varying concentrations of peptide substrates (5,000 μM). Kinetic parameters were calculated by fitting data to the Michaelis–Menten equation.

ACKNOWLEDGMENTS. We thank Nicole King for advice and for spearheading the Monosiga genome project and its use as a model organism and Anne Ashley for superb graphical skills and responsiveness. This work was supported by National Institutes of Health Grants 1 R01 HG00164-01 and P30 CA014195 and by the Razavi Newman Center for Bioinformatics. All data and additional analysis are freely available through KinBase: http://kinbase.org/kinbase/.

Supporting Information

Manning et al. 10.1073/pnas.0801314105

SI Text

RM1 Motif. This Monosiga-specific motif of ~22 aa is repeated 8–13 times in the extracellular regions of two RTKA kinases, one RTKG kinase, and 40 other Monosiga predicted proteins (see kinase.com for sequences and domain analysis). Fourteen of those proteins have N-terminal predicted signal peptides, but none have likely transmembrane regions, and only one has additional known domains (EF hands). While these gene predictions are preliminary, this suggests that some of these proteins might be secreted and possibly interact with the RTKs via homophilic adhesion.

No clear examples of the domain repeat are found outside of Monosiga, though there are some scattered weakly similar sequences particularly in bacterial surface proteins, and profile-profile analysis with pre shows a weak overlap between RM1 and part of the eukaryotic Recep.L domain. The logo view (Fig. S4) below of the alignment of all Monosiga RM1 motifs shows a partially conserved LxxL repeated pattern within the motif, which appears to be the main feature shared in these weak hits.

RM2. This ~80-aa domain is found in the cytoplasmic tail of four of the nine RTKB kinases and is repeated six times in RTKB2. It has not been found elsewhere in Monosiga or any published sequence. There is some substructure within the domain, including four conserved tyrosines followed by acidic residues (Fig. S5). These score highly by Scansite prediction (http://scansite.mit.edu) both as Src phosphorylation sites and SH2 binding sites. This domain overlaps the MR motif seen in RTKB2, but due to the substructure within the domain, the MR phase is different to that of RM2.

RM10-LRR. The RM10 motif emerged from a MEME search, and is found to partially overlap with the Pfam LRR (leucine-rich repeat) domain, so appears to be a Monosiga-specific extension of that domain. “LRR-RM10” annotations refer to the merged domain.

LDLRa. Similar to RM10, we found a variant LDL receptor type A repeat using Smart and Pfam models, and extended with a Monosiga-specific sequence extension. Unlike many other proteins, this domain is found only once per gene, and is specific both to Monosiga and to RTKs.

HYR-Related Domains. A number of weakly scoring HYR (Hyalin Repeat) domain hits resolved into three major subclasses of this domain (HYR2, HYR3, HYR4), with distinct patterns of conservation within the domain, but also considerable sequence variation, indels and partial hits within each domain, so this classification should be used with caution. HYR2 domains are most common in kinases, while HYR3 is found predominantly in SH2 proteins.

Fig S1.  Domain architecture of all Monosiga TKs.
Receptor Tyrosine Kinases

Fig S1. Continued.
Receptor Tyrosine Kinases

RTKG1
RTKG2
RTKH1
RTKH2
RTKJ1
RTKJ2
RTKK1
RTKK2

Unclassified Tyrosine Kinases

UTK01
UTK02
UTK03
UTK04
UTK05
UTK06
UTK07
UTK08
UTK09
UTK10
UTK11
UTK12
UTK13
UTK14
UTK15
UTK16
UTK17
UTK18
UTK19
UTK20
UTK21
UTK22
UTK23
UTK24
UTK25
UTK26

Fig S1. Continued.
Fig S2. Domain architecture for all Monosiga PTP, SH2 and PTB domain containing proteins. SH2 domains in kinases and PTPs are listed under those headings.
**SH2: Orthologs**

Small GTPase Signaling

- 33980 (Vav)
- 31066 (Vav)
- N7 (Vav)
- N3 (RasA1)
- 23795 (Rin)

Adaptor

- 25437 (Crk)
- 25438 (Crk)
- N6 (Grb2)

Chromatin

- 34162 (Bupf60)

Cytoskeleton

- 12682 (Tensin)

Phospholipid Signaling

- 27062 (Ship2)
- N22 (PLCγ)
- N4 (PK3R)

Scaffold

- 27493 (Shc)

Transcription

- 38850 (STAT)

Ubiquitination

- 36983 (Chb)

Unknown

- 30425 (SH2D4)

---

**SH2: Myosins**

- 8842
- 25921
- 32425
- 34161
- N20
- N5
- N8

---

Fig S2. Continued.
Fig S2. Continued.
Fig S2. Continued.
Fig S3. HMM logo comparison of *Monosiga* TKs with those of human, *Drosophila*, and *C. elegans.*
Fig S4. Logo view of RM1 motif.
Fig S5. Logo view of RM2 motif.
### Table S1. Accessory domain and motifs in *Monosiga* TKs

<table>
<thead>
<tr>
<th>Name</th>
<th>No. genes (families)</th>
<th>Copies/ gene</th>
<th>Related to/description</th>
<th>Human TKs with domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Extracellular motifs and domains</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RM1</td>
<td>3 (RTKA, G)</td>
<td>8–13</td>
<td>Unique to choanoflagellates</td>
<td>-</td>
</tr>
<tr>
<td>HYR</td>
<td>11 (RTKC, E)</td>
<td>3</td>
<td>Family of domains, related to Ig, FN3</td>
<td>EGFR, Trk, VEGFR, Tie, Axl, PDGFR, CCK4</td>
</tr>
<tr>
<td>LD LRL</td>
<td>125</td>
<td>1</td>
<td>Similar to part of LDL receptor A motif</td>
<td>-</td>
</tr>
<tr>
<td>Recep _L_domain</td>
<td>5 (RTKA, G, J, UTK)</td>
<td>1–2</td>
<td>Fragment of domain found in EGF, Insulin receptors</td>
<td>(Ig): FGFR, Trk, VEGFR, Tie, Axl, PDGFR, CCK4</td>
</tr>
<tr>
<td>FG-GAP</td>
<td>9 (FGTK)</td>
<td>3–20</td>
<td>Alpha-Integrin repeat motif</td>
<td>EGFR, InsR</td>
</tr>
<tr>
<td>EGF/CA-EGF</td>
<td>10 (RTKB-D, H J)</td>
<td>1–9</td>
<td>Epidermal Growth Factor repeats</td>
<td>Tie, Eph, ALK</td>
</tr>
<tr>
<td>LRR</td>
<td>11 (FGTK, LRTK, RTKE, L)</td>
<td>1–4</td>
<td>Leucine Rich Repeat</td>
<td>Trk</td>
</tr>
<tr>
<td>Cys-rich</td>
<td>21 (RTKB-E, J, M)</td>
<td></td>
<td>Rich in C and CxxC. Weakly similar to TNFR, furin, GCC2 repeats</td>
<td>(Furin) EGFR, InsR</td>
</tr>
<tr>
<td>ANF_receptor</td>
<td>2 (UTK, RTKC)</td>
<td>1</td>
<td>Ligand binding domain of RGCs, which contain an inactive kinase domain</td>
<td>RGC</td>
</tr>
<tr>
<td>FN3</td>
<td>5 (RTKC, RTKH, UTK)</td>
<td>1–2</td>
<td>Fibronectin Type 3 domain</td>
<td>Axl, Eph, InsR, Sev, Tie</td>
</tr>
<tr>
<td><strong>Intracellular motifs and domains</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SH2</td>
<td>14 (SFK, FVTK, 1 CTKA, 3 UTK)</td>
<td>1</td>
<td>Ptyr binding</td>
<td>Src, Tec, Abl, Csk, Fer, Syk</td>
</tr>
<tr>
<td>SH3</td>
<td>8 (SFK)</td>
<td>1</td>
<td>Binds PxxP motifs</td>
<td>Src, Tec, Abl, Csk</td>
</tr>
<tr>
<td>PTB</td>
<td>9 (HMTK)</td>
<td>1–4</td>
<td>Peptide and ptyr binding</td>
<td>-</td>
</tr>
<tr>
<td>FYVE</td>
<td>2 (FVTK)</td>
<td>1</td>
<td>Zinc Finger implicated in lipid binding</td>
<td>-</td>
</tr>
<tr>
<td>CAP.GLY</td>
<td>9 (RTKC)</td>
<td>1</td>
<td>Cytoskeleton-associated (19 copies in genome, including one PTP)</td>
<td>-</td>
</tr>
<tr>
<td>PH</td>
<td>2 (CTKA, Tec)</td>
<td>1</td>
<td>Binds to lipids and signaling proteins</td>
<td>Tec</td>
</tr>
<tr>
<td>CH</td>
<td>1 (CTKB)</td>
<td>1</td>
<td>Calponin Homology. Actin-binding and signaling roles, also seen in many SH2-containing proteins</td>
<td>-</td>
</tr>
<tr>
<td>C2</td>
<td>1 (Src)</td>
<td>1</td>
<td>Ca-dependent lipid association, maybe a substitution for missing myristoylation site</td>
<td>-</td>
</tr>
<tr>
<td>SAM</td>
<td>1 (UTK)</td>
<td>1</td>
<td>Sterile Alpha Motif, also seen in many SH2-containing adaptors</td>
<td>ACK</td>
</tr>
<tr>
<td>RM2 (<em>MR</em>'3))</td>
<td>4 (RTKB)</td>
<td>1–6</td>
<td>Novel motif, C-terminal of kinase domain. 3 conserved tyrosine residues include conserved Src-like phosphorylation/SH2 binding motif.</td>
<td>-</td>
</tr>
</tbody>
</table>
Other Supporting Information Files

Dataset S1 (PDF)  
Dataset S2 (XLS)