Review

Evolution of SH2 domains and phosphotyrosine signalling networks

Bernard A. Liu1 and Piers D. Nash2,*

1Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada
2Ben May Department for Cancer Research and Committee on Cancer Biology, The University of Chicago, Chicago, IL 60637, USA

Src homology 2 (SH2) domains mediate selective protein–protein interactions with tyrosine-phosphorylated proteins, and in doing so define specificity of phosphotyrosine (pTyr) signalling networks. SH2 domains and protein-tyrosine phosphatases expand alongside protein-tyrosine kinases (PTKs) to coordinate cellular and organismal complexity in the evolution of the unikont branch of the eukaryotes. Examination of conserved families of PTKs and SH2 domain proteins provides fiduciary marks that trace the evolutionary landscape for the development of complex cellular systems in the proto-metazoan and metazoan lineages. The evolutionary provenance of conserved SH2 and PTK families reveals the mechanisms by which diversity is achieved through adaptations in tissue-specific gene transcription, altered ligand binding, insertions of linear motifs and the gain or loss of domains following gene duplication. We discuss mechanisms by which pTyr-mediated signalling networks evolve through the development of novel and expanded families of SH2 domain proteins and the elaboration of connections between pTyr-signalling proteins. These changes underlie the variety of general and specific signalling networks that give rise to tissue-specific functions and increasingly complex developmental programmes. Examination of SH2 domains from an evolutionary perspective provides insight into the process by which evolutionary expansion and modification of molecular protein interaction domain proteins permits the development of novel protein-interaction networks and accommodates adaptation of signalling networks.

Keywords: SH2; protein-tyrosine kinase; eukaryotes; phosphotyrosine

1. INTRODUCTION

Phosphotyrosine (pTyr)-mediated signalling is a relatively recently evolved cellular communication system in comparison to more primordial post-translational modifications (PTMs) such as Ser/Thr phosphorylation or lysine ubiquitination that are used across all branches of the eukaryota [1]. Some prokaryotes possess unique tyrosine kinases that are involved in pTyr-mediated cellular activities but these are distinct from those found in metazoans (see Grangeasse et al. [2]). Signalling mediated through the use of pTyr is essential in metazoans, regulating many key cellular and developmental processes, including cell growth, proliferation, differentiation and migration. In metazoans, pTyr signalling depends upon an essential triad of signalling molecules: the protein-tyrosine kinases (PTKs) that add a phosphate onto substrate tyrosines (conceptually these may be viewed as the ‘writers’); the protein-tyrosine phosphatases (PTPs) that remove or dephosphorylate substrates (the ‘erasers’); and the modular protein interaction domains that recognize the phosphorylated ligand (the ‘readers’) and hence recruit the proteins containing these domains to specify downstream signalling events [1]. Several modular interaction domains have the capability to bind to tyrosine-phosphorylated ligands. These include most Src homology 2 (SH2) domains [3,4], a subset of phosphotyrosine-binding (PTB) domains [5], at least one C2 domain [6] and the Hakai pTyr-binding domain [7]. The SH2 domain is the largest domain family dedicated to pTyr recognition, with 111 proteins containing at least one SH2 domain encoded in the human genome [8]. With the recent sequencing of the genomes of multiple organisms, the identification of the PTKs, PTPs and SH2 domains has provided insights into the emergent evolution of this PTM signalling system [1,9–11]. Recent examination of the evolutionary history of SH2 domains indicates that PTKs coevolved along with SH2 domain proteins to bring about organismal and tissue complexity in metazoans [8]. The diversification of proteins that contain SH2 domains allows this singularly remarkable domain to coordinate specificity in a diverse range of signalling networks.

This review explores the origins of pTyr-mediated signalling through the expansion and diversification of SH2-domain-containing proteins. Examination of the origins of SH2 domains and their respective protein families reveals that SH2 domains are integrated into a diverse set of protein interaction modules. In this way, SH2 domains coordinate
specificity across a multitude of signalling pathways. As we trace the evolutionary history of the different protein families, we discuss the various evolutionary paths by which these proteins change and diverge. In detail, we will highlight the key underlying genetic mechanisms of diversification through gene duplication, transcriptional regulation, domain loss and gain, and insertion of linear motifs. Lastly, we will explore both the conserved and divergent nature of pTyr-mediated signalling networks to provide a better appreciation for the complexity of mammalian pTyr networks. This overview will introduce the mechanisms by which adaptive changes within pTyr–SH2 interaction networks act to drive cellular complexity and participate in the development of tissue specific signalling networks. Understanding the molecular changes that drive expansion and complexity of pTyr signalling networks provides an early glimpse of some aspects of the process of molecular evolution that underlies the adaptive changes in organisms as they adapt and develop more elaborate functions and increasingly complex tissue types.

2. SH2 DOMAINS HAVE A BROAD ROLE IN PHOSPHOTYROSINE SIGNALLING AMONG MAMMALS

Much of our knowledge and understanding of pTyr signalling and SH2 domains comes from studies in mammals, particularly in mouse and human. The human genome encodes roughly 90 PTKs, 107 PTPs, 54 PTB domain-containing proteins (only about one-fifth of PTB domains can bind pTyr) and 111 SH2-domain-containing proteins. This creates a complex network of regulators in pTyr signalling (figure 1a) [3,12,13]. Since the discovery of pTyr nearly three decades ago [14], pTyr signalling has been implicated in almost every aspect of normal cellular function and found to be a major driver in human malignancies such as cancer [15,16]. Tyrosine phosphorylation encompasses a small percentage (up to 2 per cent) of the total protein phosphorylation within a mammalian cell, although a list of over 10 000 putative pTyr sites have been identified so far across various tissues and species [17,18]. The discovery of the SH2 domain [19–21] led to numerous
studies mapping interactions and SH2 domain selectivity that have led to a growing understanding of specificity of pTyr signalling (reviewed in earlier studies [3,4,22]). SH2 and PTB domains determine selectivity for pTyr ligands by recognizing pTyr and residues surrounding the pTyr (figure 1a). The preference for +1 to +5 residues C-terminus of the pTyr in a sequence-dependent manner is dictated by the surface residues on the SH2 domain that lie adjacent to the pTyr-binding pocket. In contrast, PTB domains recognize sequences N-terminus to the pTyr with a core consensus of N-P-x-pY, yet only a small fraction of PTB domains are capable of binding pTyr peptides. An overwhelming preponderance of what we know about pTyr signalling is based on studies in human and mouse systems, with very little known about pTyr signalling in simpler organisms. Mammalian genomes and protein interaction maps provide one window into the role of pTyr signalling in the development of complex organisms. Indeed, a study of human PTK circuits using prediction networks to identify PTKs, SH2 and PTPs suggests an evolving paradigm of pTyr signalling and evolution [23]. In this review, we describe the evolutionary provenance and trajectory of aspects of the cellular pTyr circuitry. The available pTyr machinery in various extant organisms suggests that pTyr use transitioned from a nuclear and cytoplasmic system in early unicellular eukaryotes to signalling at the cellular membrane through use of receptor tyrosine kinases (RTKs). This evolutionary analysis of pTyr signalling has implications for the involvement of PTKs and pTyr signalling in human disease such as cancer [23].

The spatio-temporal regulation of tyrosine phosphorylation is necessarily complex in mammalian cells. The interplay between the substrate tyrosines and the array of available PTKs and PTPs governs the timing, localization and longevity of phosphorylation events. PTKs and PTPs commonly have somewhat limited substrate specificity for peptide ligands. They are, however, tightly controlled in both their activation and spatial localization. The result is that, as a general paradigm, PTKs and PTPs control the spatio-temporal localization of phosphorylation events but have limited deterministic control over downstream events. Specificity in pTyr signalling is thus largely determined by recognition modules, such as SH2 domains, that distinguish specific tyrosine phosphorylated motifs and act to nucleate complex formation to propagate signalling. Further complexity arises from tissue- and cell-specific expression profiles and dynamic sub-cellular localization of reader proteins that may play a role in the cellular context of protein phosphorylation. Therefore, dissecting the temporal and spatial organization of dynamic protein interaction networks is an essential element in understanding information flow through the signalling system [24].

Transgenic animal models give a first approximation of the role of a protein in a signalling network. Over 80 of the 111 SH2 proteins have been genetically disrupted in mice, revealing a broad range of cellular functions for SH2 domains in various tissues and cell types [3]. However, in order to extrapolate the multiple biological functions of these proteins and understand the cellular context by which these proteins regulate cellular activity, understanding of the PTKs, PTPs and potential interaction networks within specific cell types or tissues is required. This already challenging problem is compounded by the copious variety of tissues and cell states. Further complexity arises when multiple members of a closely related gene family are present where some extent of overlapping or redundant functions coexist with specialized roles. For instance, in the GRB2 family (Grb2, Gads and Grap), Grb2 is ubiquitously expressed and plays an essential role in growth factor signalling downstream of RTKs. Gads and Grap, by contrast, are highly restricted in their expression to a subset of immune cells and appear to function primarily in signalling downstream of the T-cell and B-cell receptors. Loss of Grb2 results in embryonic lethality [25], whereas loss of Gads or Grap leads to rather specific immune phenotypes [26,27]. While the three encoded genes are closely related, they possess distinct roles in their specific cellular context. This is not always the case for all families, suggesting that some members have redundant roles while some individual genes within a family have a ubiquitous and essential role in development [3]. We discuss some of the evolutionary patterns that drive divergence and complexity in SH2-domain-containing proteins in §6.

3. IDENTIFICATION AND CLASSIFICATION OF SH2 DOMAINS

Large numbers of whole genome sequences, combined with emerging bioinformatics tools, present the opportunity to examine signalling systems across a broad range of organisms. Modular domain families are identified from sequence datasets using hidden Markov models and multiple sequence alignments tools such as those developed by the protein families database (Pfam) [28], simple modular architecture tool (SMART) [29] and conserved domains database (CDD) [30]. Soon after the availability of their annotated genomes, the human and mouse complements of all the kinases, including the PTKs, as well as the PTPs, were reported [12,13,31,32]. This was followed by the identification of 111 SH2 domain proteins in the human and mouse genomes [3] and later by the total complement of PTKs and SH2 encoding genes across multiple eukaryotic organisms (figure 1b) [8,10,11]. The 111 SH2-containing proteins contain a diverse array of additional protein interaction domains (e.g. SH3, PTB, PH) and catalytic domains (e.g. TK, phosphatase) that are gained and lost over time to generate diversity in signalling functions [3,33]. With over 150 eukaryotic genomes sequenced till date, bioinformatic approaches are required to identify protein domains and trace paralogues and orthologues across these diverse organisms. By examination of SH2 domains across 21 extant species in the Eukaryota we identified 38 conserved families of SH2 domains using a hierarchical method combining sequence homology, domain organization and positions of splice-site patterns [3,8]. Such a hierarchical method was necessary to avoid improper classification of a domain family based on sequence homology alone. Analysis of splice-site position within the open reading frame provides an easily accessible hallmark.
of conserved genomic structure that can assist in accurate classification of a family [8,34]. This provided a methodology that could trace families as well as individual proteins across lineages and species. Domain families assist in defining orthologous and paralogous proteins across the 21 different eukaryotes. This in turn provides insights into the evolutionary mechanisms that drive diversification within these families that create new proteins with both shared and novel functions.

The 38 families of SH2 domains in humans encode a heterogeneous range of proteins that connect pTyr signalling to a diverse array of cellular systems. Many families have distinctive domain architectures that place the SH2 domain in the context of other modular protein domains such as SH3, PH, kinase, phosphatase, RING, PTB, RasGAP, RasGEF, FERM and others (table 1) [3,33]. The various mechanisms by which diversity appears to have arisen within individual families will be examined in detail in later sections. Examination of these families across a broad range of eukaryotes suggests points of origin for each SH2 family at common ancestors representing branch points that separate specific lineages. The role of each SH2 protein and selective advantage conferred upon first emergence is rarely clear and presents a major challenge for understanding the core cellular functions of primordial SH2 domain proteins. Sequence analysis and functional studies provide a means to trace the evolutionary history of SH2-mediated pTyr signalling in metazoans. Examination of the coevolution of PTKs and SH2 domain proteins may provide insights into their role in the development of evolutionary mechanisms that drive diversification within these families that create new proteins with both shared and novel functions.

### Table 1. SH2 domain families.

<table>
<thead>
<tr>
<th>family</th>
<th>human gene names</th>
<th>domains (in order)</th>
<th>pseudogenes (yes/no)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL</td>
<td>Abl1, Abl2</td>
<td>SH3, SH2, TK , F-actin binding domain</td>
<td>no</td>
</tr>
<tr>
<td>CBL</td>
<td>Cbl, CblB, CblC</td>
<td>EF-Hand, SH2, 4-helix bundle, RING, UBA*</td>
<td>no</td>
</tr>
<tr>
<td>CHN</td>
<td>Chn1, Chn2</td>
<td>SH2, C1, RhoGAP</td>
<td>no</td>
</tr>
<tr>
<td>CRK</td>
<td>Crk, CrkL</td>
<td>SH2, SH3, SH3</td>
<td>no</td>
</tr>
<tr>
<td>CSK</td>
<td>Csk, Math</td>
<td>SH3, SH2, TK</td>
<td>no</td>
</tr>
<tr>
<td>DAPPI</td>
<td>Dapl</td>
<td>SH2, PH</td>
<td>no</td>
</tr>
<tr>
<td>FES</td>
<td>Fes, Fer</td>
<td>F-Bar, SH2, TK</td>
<td>no</td>
</tr>
<tr>
<td>FRK</td>
<td>Frk, Brk, Srms</td>
<td>SH3, SH2, TK</td>
<td>no</td>
</tr>
<tr>
<td>GRB2</td>
<td>Grb2, Gads, Grap</td>
<td>SH3, SH2, SH3</td>
<td>no</td>
</tr>
<tr>
<td>GRB7</td>
<td>Grb7, Grb10, Grb14</td>
<td>RA, PH, SH2</td>
<td>no</td>
</tr>
<tr>
<td>JAK</td>
<td>Tyk2, Jak1, Jak2, Jak3</td>
<td>FERM, SH2, TK</td>
<td>no</td>
</tr>
<tr>
<td>NCK</td>
<td>Nck1, Nck2</td>
<td>SH3, SH3, SH3, SH2</td>
<td>no</td>
</tr>
<tr>
<td>PI3KR</td>
<td>Pih3r1, Pih3r2, Pih3r3</td>
<td>SH3*, RhoGAP*, SH2, SH2</td>
<td>no</td>
</tr>
<tr>
<td>PLCγ</td>
<td>Pieg1, Pieg2</td>
<td>PH, PLCXc, PH, SH2, SH2, SH3, PLCYc, C2</td>
<td>no</td>
</tr>
<tr>
<td>PTPN</td>
<td>Ppmp6, Ppmp11</td>
<td>SH2, SH2, phosphatase</td>
<td>yes</td>
</tr>
<tr>
<td>RASA1</td>
<td>Ras1</td>
<td>SH2, SH3, SH2, PH, C2, RasGAP</td>
<td>no</td>
</tr>
<tr>
<td>RIN</td>
<td>Rin1, Rin2, Rin3</td>
<td>SH2, VPS9, RA</td>
<td>no</td>
</tr>
<tr>
<td>SH2B</td>
<td>Apn, Lnk, Sh2b</td>
<td>PH, SH2</td>
<td>no</td>
</tr>
<tr>
<td>SH2D1</td>
<td>Sh2d1a, Sh2d1b</td>
<td>SH2</td>
<td>yes</td>
</tr>
<tr>
<td>SH2D2</td>
<td>Sh2d2a, Sh2d2b, Sh2d7</td>
<td>SH2</td>
<td>no</td>
</tr>
<tr>
<td>SH2D3</td>
<td>Sh2d3a, Sh2d3b, Bcar3</td>
<td>SH2, RasGEF</td>
<td>no</td>
</tr>
<tr>
<td>SH2D4</td>
<td>Sh2d4a, Sh2d4b</td>
<td>SH2</td>
<td>no</td>
</tr>
<tr>
<td>SH2D5</td>
<td>Sh2d5</td>
<td>PTB, SH2</td>
<td>no</td>
</tr>
<tr>
<td>SH3BP2</td>
<td>Sh3bp1, Sh3bp2</td>
<td>PH, SH2</td>
<td>no</td>
</tr>
<tr>
<td>SHB</td>
<td>Shb, Shd, Shc, Shf</td>
<td>SH2</td>
<td>no</td>
</tr>
<tr>
<td>SHC</td>
<td>Shc1, Shc2, Shc3, Shc4</td>
<td>PTB, SH2</td>
<td>yes</td>
</tr>
<tr>
<td>SHIP</td>
<td>Ship1, Ship2</td>
<td>SH2, phosphatase</td>
<td>no</td>
</tr>
<tr>
<td>SLAP</td>
<td>Slap, Slap2</td>
<td>SH3, SH2</td>
<td>no</td>
</tr>
<tr>
<td>SLPT6</td>
<td>Slp6, Blnk, Slp76, Mist</td>
<td>SAM*, SH2</td>
<td>no</td>
</tr>
<tr>
<td>SOCS</td>
<td>Socs1, Socs2, Socs3, Socs4, Socs5, Socs6, Socs7, Cish</td>
<td>SH2, Socs</td>
<td>yes</td>
</tr>
<tr>
<td>SRC</td>
<td>Src, Lyn, Lck, Fgr, Yes, Lyn, Hck, Blk</td>
<td>SH3, SH2, TK</td>
<td>yes</td>
</tr>
<tr>
<td>STAP</td>
<td>Bks, Brad1</td>
<td>PH, SH2</td>
<td>no</td>
</tr>
<tr>
<td>STAT</td>
<td>Stat1, Stat2, Stat3, Stat4, Stat5a, Stat5b, Stat6</td>
<td>DNA BD, SH2, TA</td>
<td>no</td>
</tr>
<tr>
<td>SPT6</td>
<td>Supt6h</td>
<td>CSZ, S1, SH2</td>
<td>no</td>
</tr>
<tr>
<td>SYK</td>
<td>Zap70, Syk</td>
<td>SH2, SH2, TK</td>
<td>no</td>
</tr>
<tr>
<td>TEC</td>
<td>Bmx, Tec, Btk, Irb, Ttk</td>
<td>PH, BTK, SH3, SH2, TK</td>
<td>no</td>
</tr>
<tr>
<td>TNS</td>
<td>Twn1, Twn3, Twn1c, Twn4</td>
<td>PTPc, SH2, PTB</td>
<td>no</td>
</tr>
<tr>
<td>VAV</td>
<td>Vav1, Vav2, Vav3</td>
<td>CH, RhoGEP, PH, C1, SH3, SH2, SH3</td>
<td>no</td>
</tr>
</tbody>
</table>

*Human pseudogenes identified using the Pseudofam database (http://pseudofam.pseudogene.org/) [35]; **domains that are not shared among all family members. For complete domain descriptions visit http://pawsonlab.mshri.on.ca or http://www.sh2domain.org.
multicellularity and subsequent leaps in complexity of tissues and development of complex systems such as the immune system and the mammalian brain (see §5).

4. SH2 DOMAINS PREDATE DEDICATED PROTEIN-TYROSINE KINASES

Within Eukaryota, there exist two major divisions: the bikonts (eukaryotic organisms derived from cells with two flagella) and unikonts (organisms derived from cells with one flagella). SH2 domains were previously described in unikont organisms such as humans, yeast, amoeba and choanoflagellates [10,11,36]. Recent analysis of SH2 domains across 21 eukaryotic genomes revealed the presence of SH2 domains in both divisions [8], including the unicellular organisms Naegleria gruberi and Phytophthora capsici, and thus SH2 domains probably predate the emergence of PTKs [37]. A common SH2 domain protein, Spt6, is found in both bikonts and unikonts and is universally present in metazoans. A handful of other SH2 domains are observed in certain bikont lineages, but it remains unclear whether any of these are functional and capable of binding pTyr [8]. Spt6 was first described in the budding yeast Saccharomyces cerevisiae [36] as an elongation factor, a histone H3–H4 chaperone that mediates nucleosome assembly and interacts with RNA polymerase II (RNAPII) [38,39]. The SH2 domain of Spt6 is reportedly capable of recognizing the phosphorylated serines (pSer) of RNAPII C-terminal tail (CTD) [40], with no detectable binding to pTyr peptides. Structural studies indicate that Spt6 possesses a tandem SH2 fold [41,42] (figure 2a). Only the first SH2 domain of Spt6 is recognizable as a canonical SH2 domain by sequence analysis. The second SH2 fold is highly atypical, lacking the FLVR (Phe-Leu-Val-Arg) motif for phospho-binding, and is not detected using standard SH2 domain models, such as SMART or Pfam. Both SH2 domains are necessary for this region to recognize pSer residues on the CTD of RNAPII [40]. The structure of the yeast Spt6 N-SH2 domain possesses many sequence and structural features that closely resemble pTyr-binding SH2 domains in mammals [41,43–45]. Binding to pSer/pThr peptides uses the same canonical binding pocket as canonical SH2 domains use to bind pTyr peptides [45]. Based on sequence and structure, it appears likely that the SH2 domain of Spt6 represents an ancestral pSer/pThr-binding domain that may have given rise to pTyr-binding SH2 domains.

Figure 2. Primordial function of SH2 domains and pTyr signalling. (a) The ribbon structure of the yeast Spt6 tandem SH2 (left panel, PDB: 3PJP). The charged surface of the tandem SH2 (right) indicates the positions of negative (red) and positive charge (blue) of the tandem SH2 domain. Highlighted in the yellow circle is the phospho-binding pocket for pSer. (b) The SH2 domain transcriptional activator STAT (red) is present as early as Dictyostelium and found in humans. Analogous to mammalian STAT signalling, STATc is regulated by tyrosine phosphorylation (yellow dot) through PTKs (green) and PTPs (blue). In Dictyostelium, the SH2/RING domain containing protein CblA (orange) antagonizes the phosphastase, while in mammals Cbl inhibits the PTK Jak.
somewhere around the divergence between Unikonta and Bikonta. The Spt6 SH2 remains well conserved from yeast to humans.

Most branches of Unikonta, including Amoebozoa, Choanozoa and Animalia, contain diverse sets of SH2 domains, suggesting that pTyr-binding SH2 domains developed very early in the unikont lineage. In members of the Amoebozoa lineage, such as the Entamoebae and Mycetozoa, SH2 domains expand into an independent set of proteins with relatively few obvious homologues in Metazoa (e.g. Shk, LrrB) [46,47]. This is despite the absence of dedicated PTKs that existed within these species. It is not immediately clear how pTyr signalling networks are constructed within these organisms. Some Ser/Thr kinases, such as dual specificity MAPks [48], as well as other unique PTKs such as DPYK [49], are capable of phosphorylating tyrosine residues. Such non-traditional tyrosine kinases may explain the expansion and maintenance of SH2 domain proteins in Amoebozoa such as Dictyostelium discoideum. It is also in Amoebozoa that we first encounter SH2 domains linked to a dual-specificity kinase, the Ser/Thr/Tyr kinase Shk [50]. Despite the absence of PTKs, tyrosine phosphorylation has been reported in D. discoideum and can be observed in the phosphorylation of the C-terminus of Statc [51,52] (figure 2b). When Statc is tyrosine phosphorylated, it forms a homodimer through reciprocal SH2-domain–pTyr interactions and then accumulates in the nucleus, functioning in a manner analogous to the mammalian STAT orthologues. Amoebozoa, Choanozoa and Metazoa all encode SH2-domain-containing Stat proteins. In addition, D. discoideum contains a Cbl-like pTyr-binding region, comprising an EF-hand and an SH2 domain, linked to an E3 ubiquitin-ligase RING domain. D. discoideum CblA functions as a positive regulator of Statc by downregulating a tyrosine phosphatase, whereas metazoan CBL proteins negatively regulate tyrosine kinases [52]. Because D. discoideum diverged from the lineage that gave rise to Metazoa before Fungi but after the Bikonta/Unikonta split, it remains unclear why Stat, CBL and other SH2 domain proteins were apparently lost in Fungi [53]. In addition to the SH2 domain proteins, 24 kinase subfamilies shared between D. discoideum and Metazoa are absent in yeast, suggesting that yeast developed a specialized biological programme that does not require many of the kinase signalling networks used in other unikont branches [54]. In fact, the progenitor to the yeasts may have used pTyr signalling, but, as Gerard Manning's analysis suggests, the PTKs and some other kinase families were lost as the yeasts evolved leaving a few PTP/DSPs (dual specificity phosphatases). Fungi, which together with Metazoa are in the opisthokont branch of Unikonta, lack any apparent pTyr-binding SH2 domains other than the pSer-binding SPT6. pTyr-binding SH2 domains were lost from Fungi, perhaps as a result of deleterious outcomes from the occurrence of pTyr. This is supported by lethal effects of expression of the PTK v-Src in S. cerevisiae that are associated with loss of cell cycle control [55,56]. The ability of v-Src to kill yeast cells is dependent upon its SH2 domain as an active kinase alone has little effect on yeast growth [57,58]. This suggests that pTyr-mediated signalling may have been negatively selected out of the Fungi common ancestor around 1200 Ma [59,60] and that functional use of pTyr and SH2 domains may predate dedicated PTKs [47].

Tyrosine phosphorylation and tyrosine kinase signalling is thought to occur primarily in the cytoplasm, particularly at the plasma membrane, in metazoans. Upon ligand engagement or cell contact, activated RTKs phosphorylate their own cytoplasmic tails as well as membrane-localized substrate proteins. SH2 domain proteins are then recruited to activated signalling complexes at the plasma membrane or on signalling endosomes to propagate downstream signalling. In this paradigm, much of the PTK–SH2 signalling is cytoplasmic in nature with only limited examples of nuclear involvement (however, more and more nuclear pTyr proteins are being reported, although admittedly this does not necessarily tell us where in the cells these proteins were phosphorylated or whether they have nuclear SH2-domain-binding partners). In contrast, primordial SH2 domain-containing proteins, SPT6 and Stat, function in the nuclear compartment of the cell. There was probably significant nuclear pTyr signalling in early unicellular eukaryotes that remains a feature of various lineages. This is supported by several lines of evidence that tyrosine phosphorylation plays a role in the nucleus. Tyrosine phosphorylation of nuclear cyclin cycle regulators such as Cdc28 and Cdc2 is observed in mammalian and yeast cells [61–63]. Staining for pTyr in the unicellular organism Giardia lamblia indicates enrichment in the nuclei that is distinct in comparison to pSer/Thr staining [64]. This suggests that in early unicellular eukaryotes, pTyr signalling may have been more central to the nuclei with concurrent evolution of SH2 domains that may mediate nuclear functions such as DNA transcription (e.g. Stat proteins). As cytoplasmic tyrosine kinases (CTKs) emerged, pTyr signalling may have become more extensively used for cytoplasmic functions, eventually becoming focused at the plasma membrane with the emergence of RTKs. Extensive RTK signalling mediated by SH2 domains is itself a hallmark of multicellularity in Metazoa and represents a key stepping stone in the emergence of multicellularity in the animal lineage [23].

Early in the unikont lineage, SH2 domains appear that have the ability to recognize pTyr-based motifs. This is a key feature allowing rapid evolution of pTyr signalling by shuffling SH2 domains into various proteins [1]. In pre-metazoans, specifically the choanoflagellate lineage, diversification of SH2 domains and PTKs is tightly linked, suggesting that they coevolved to create diverse functional signalling systems. Within the unikont lineages, SH2 domain expansion may be partially driven by the kinases that generate the binding partners for SH2 domains through phosphorylation. True PTKs first appeared in the unicellular choanoflagellate Monosiga brevicolis [65], but are absent in amoeba and slime moulds [54]. The choanoflagellate branch, which includes M. brevicolis, encodes the core set of PTKs, PTPs and SH2 domains [10,11] and thus exhibits a functional pTyr signalling
network reminiscent of metazoans. A total of 19 conserved SH2 families and six conserved PTK families are found in *M. brevicollis*, several of which contain both an SH2 and a TK domain (figure 3). The heterogeneity of the *M. brevicollis* pTyr system suggests that SH2 proteins may have cellular roles that diverge somewhat from their mammalian counterparts. For instance, the CTK MbCsk is co-expressed with and can phosphorylate MbSrc1 at tyrosine near the C-terminus in a manner similar to that of the mammalian Src and Csk pair. However, the negative regulation of Src through phosphorylation of this residue is absent from the *M. brevicollis* pair, suggesting that allosteric regulation of Src developed more recently in the metazoan lineage [66,67]. The presence of an extensive and varied collection of pTyr signalling molecules suggests that *M. brevicollis* depends upon pTyr to mediate a multitude of cellular functions, most of which we have yet to understand. It will be fascinating to see what roles unique SH2 proteins in *M. brevicollis* serve in coordinating organism-specific pTyr signalling.

5. EMERGENCE OF PROTEIN-TYROSINE KINASE AND SH2 DOMAIN FAMILIES FOR CELLULAR COMPLEXITY

Throughout the evolution of Metazoa, new SH2 and PTK families appear while others are occasionally lost (figure 4). The timing of the appearance of new families may indicate the evolution of specific signalling networks that support an adaptive advantage. Phenotypic complexity and complex developmental programmes are hallmarks of metazoan evolution. Numerous SH2 families emerge in relatively simple eukaryotes such as Choanozoa and Cnidaria. It remains unclear what the specific cellular functions

**Figure 3.** The gain and loss of SH2 and PTK families in Eukaryota. The eukaryotic tree is shown with the phylum and genus species listed on the right. The number of SH2/PTK families are listed in black text with the gain and loss indicated in grey text. The 38 SH2 and 29 PTK families represent those that are found in humans and are present in other lower eukaryotes [8,10,34].
these proteins serve in primordial cellular signalling. Certain SH2 families arise at distinct stages coincident with the development of multicellularity. For example, establishment of cell polarity in cnidarians probably depended upon the appearance of specific signalling pathways such as the canonical Wnt and β-catenin.

Figure 4. Emergence and expansion of SH2 and PTK families. The presence and absence of conserved families of SH2 and PTK proteins is indicated for representative species covering the unikont branch from social amoeba/slime mould Dictyostelium species to H. sapiens. The families of SH2 and PTK were identified previously [8,10,34]. Text colour indicates SH2 families that lack a TK domain (black), RTK families (orange), CTKs containing an SH2 domain (blue) and CTKs that lack an SH2 domain (green). The general domain organization of each family is shown on the right.
pathway [68,69]. Alongside appeared several SH2 domain families, including NCK and SOCS. Both Nck and Socs7 play an important role in regulating cell polarity through Septins in mammals [70], which suggests that they may have coevolved with the polarity complexes that emerge in cnidarians [71]. Further hints of the emerging programmes guiding multicellularity can be seen in the families of PTKs that are identified in the cnidian Nematostella vectensis. Only two families of RTKs are present in M. brevicollis, whereas nine additional families arise in N. vectensis. This correlates with studies by Li et al. [23] suggesting that early pTyr signalling was primarily mediated by CTK. In cnidarians, additional RTK families (RYK, ROR, VEGFR and others) promote inter-cellular communication for cell–cell contact and cell adhesion (figure 4). The coincident emergence of FAK, a PTK crucially involved in regulating focal adhesions critical for maintaining cell contact [72,73], during this lineage further supports the development of this complex cellular process.

In arthropods, complex tissues and organs begin to emerge, such as the muscle, heart and vascular system as well as the innate immune system [74]. The emergence of the muscle-specific tyrosine kinase in arthropods is coincident with the development of a muscular system within this lineage (figure 4). In addition to the emergence of new proteins and protein families, protein interaction networks can be seen to be coevolving in parallel. For example, Shc and Grb2 are both observed in N. vectensis (figure 4). An interaction between Shc and Grb2 is first seen in arthropods with the insertion of a pTyr site within the CH1 region of Shc [8]. Phosphorylation of this site allows for interaction with the SH2 domain of Grb2. In mammals, this interaction is supported by multiple pTyr sites in the products of four distinct Shc genes (Shc1, 2, 3, 4) expressing various isoforms in a wide variety of tissues [75]. Furthermore, elimination of these pTyr sites in the Shc1 gene in mice leads to a muscle motor stretch reflex defect, specifically a decreased number of muscle spindles of the skeletal muscle sensory organs that regulate motor behaviour [76]. This highlights the various processes for emerging novel protein circuits through evolving not just novel protein families but also bringing together the pTyr circuit through evolving pTyr motifs and SH2 domains in a manner that may have promoted the development of tissue-specialized signalling. Additional analysis regarding evolution of pTyr signalling networks will be discussed in a later section.

Although many proteins emerge at distinct points in evolution, mutation, adaptation and selection underlie a process capable of developing new interactions that drive the creation of novel signalling networks. Functional interactions in one species do not, therefore, necessarily represent primordial protein functions. Studies in mammalian systems suggest that Grb2 associates with Shc and epidermal growth factor receptor (EGFR), whereas Crk interacts with the CTK c-Abl and thus plays essential roles in these signalling networks [77–79]. Yet, these functions may be later adaptations and signify the role that these proteins served in common ancestor species or the role that they serve in other lineages. Both Grb2 and Shc appear in M. brevicollis and are found across all metazoans, suggesting that they emerged in a common pre-metazoan ancestor species (figure 4). Grb2 associates with Shc through a pTyr-dependent interaction dependent upon the SH2 domain of Grb2 and the pY-x-N motifs present in the CH1 region of Shc. However, these pTyr-binding sites do not appear until arthropods such as Drosophila melanogaster, suggesting that the ancestral role for Grb2 and Shc were independent of this function, and that this is a later adaptation. Phenotypic evidence supports the assertion that the Grb2–Shc–pTyr-mediated interaction is indeed functionally important in signalling events in more complex multicellular metazoans and in particular for signalling networks that underlie the development of neuromuscular systems. In a similar manner, the Jak-Stat signalling network is extensively used in adaptive immune signalling in mammals [80] and gene disruptions in this network often exhibit specific defects in immunological signalling. Yet, both Jak and Stat proteins predate the adaptive immune system and appear to have arisen at different points in evolution. This is consistent with evolution proceeding by rewiring existing components when adaptive pressure drives the development of more complex systems. Thus, the ancestral function of proteins can either change or become obscured by observed functions in well-studied mammalian organisms. Shc1, for instance, serves essential functions in mammals distinct from the pTyr-mediated interaction with Grb2. Targeted disruption of the Shc1 gene is embryonic lethal [81], while disruption of the pTyr sites that allow Grb2 binding has a relatively specific function later in development [76]. The essential function of Shc1 may well relate to the ancestral function of the protein in single-celled pre-metazoan species. Evolution in this case has resulted in the use of existing components to build new circuits and connections that underlie novel signalling pathways and add additional levels of complexity.

6. DIVERSIFICATION OF SH2 DOMAIN PROTEINS THROUGH GENE DUPLICATION

Over 30 years ago, Ohno [82] elucidated the potential for gene duplication to drive diversification, but it was not until recently that this concept has become widespread and pivotal in understanding the evolution of gene families and functional diversification. Approximately 65 per cent of human genes are duplicated, with 33 of 38 SH2 and 25 of 29 PTK families possessing duplicate copies. Gene duplication provides the opportunity for the duplicate gene to explore an otherwise prohibited mutational space. Freed from the pressure to maintain essential function the newly duplicate gene is free to wander, on condition that the duplicate genes collectively continue to contribute their original function. In the case of SH2 domain proteins, very few new families arise at the split between arthropods and vertebrates (figure 3). Instead, gene duplication appears to function by providing an opportunity to build complexity out of existing components. The expansion and divergence of SH2 domains and pTyr signalling components

Phil. Trans. R. Soc. B (2012)
provides a rich case study for how organisms and genomes gain complexity by gene duplication followed by functional divergence [83].

To understand how gene duplicates give rise to diversification, let us consider the general mechanisms and consequences. When a gene is duplicated, it faces a number of hurdles that must be overcome for it to be fixed into the genome of a species. If a duplicate gene confers no selective advantage, it is likely to be lost or mutated. In order for a gene duplicate to be observed in an extant species, it must be first fixed within the population and second it must be preserved over time [83,84] (figure 5a). Three fates are typically envisaged for a gene duplicate. One fate, called non-functionalization, refers to the loss of all functionality of a gene. This typically leaves behind a pseudogene wherein the duplicate gene accumulates mutations that destroy its function to the point at which it can no longer be transcribed or translated. The SH2-homology pseudogene Grb2-ps1 is present within the mouse genome but is absent in humans, yet no transcript can be identified and the sequence is recognizably that of a pseudogene. At least 16 human SH2 pseudogenes can be identified using the Pseudofam database (http://pseudofam.pseudogene.org/) [35] (table 1). A second fate results when an advantageous allele arises as a result of one of the gene copies gaining a new function. This is referred to as neofunctionalization. For example, the SLAP family appears to have originated from a duplication of a SRC family gene. The loss of the TK domain in SLAP allowed this duplicate to acquire a new function distinct from its ancestor. Similarly, SAP (SH2D1A), arose from a duplication of SHIP, and lacks the phosphatase domain. Neofunctionalization is more likely to occur in genes that are necessarily rapid evolving such as those involved in immunity and host defence [85]. Lastly, the original functions of the single-copy gene may be partitioned between the duplicates. This is known as subfunctionalization. Biological function suggests that many gene products perform a multiplicity of subtly distinct functions and thus selective

Figure 5. Fates of duplicate genes. (a) When a new duplicate arises it is eliminated in a majority of cases as fixation in a population is rare event. After fixation, the gene can undergo three general fates that include: neofunctionalization, where a duplicate copy acquires a novel function; non-functionalization of a gene through mutation of its promoter and/or mutation of its function creating a pseudogene by inactivation; and subfunctionalization that divides the primary ancestral function between the duplicate copies possibly across distinct tissue types. (b) Duplicate copies of genes can diverge through means of evolving novel protein domain specificity for ligands, gain or loss of a protein domain, insertion or deletion of short linear motifs and lastly an alteration in the promoter of the gene allowing for specialized tissue expression and transcription control.
pressures may result in a compromise between optimal sequences for each role. Partitioning these functions between the duplicates may increase the fitness of the organism by removing the conflict between two or more functions. For example, the families of NCK and VAV each demonstrate subfunctionalization. Loss of either Nck1 or Nck2 in mice result in only subtle phenotypic changes [86–88], whereas loss of both alleles results in embryonic lethality. This is indicative of overlapping and mutually compensatory functions [89]. In a similar vein, loss of individual genes within the VAV family causes no obvious phenotypes, whereas loss of all members of the family results in a severe defect in lymphocyte development [90]. Subfunctionalization can also enable temporal and spatial partitioning of expression, protein localization and assembly of signalling complexes.

The driving force during evolution for preserving a duplicate of an existing gene is to gain a selective advantage. Gene duplicates must confer an immediate selective advantage (or at least not a selective disadvantage) and this is clearly the case if elevating the level of protein expression (dosage) is advantageous. The dosage balance involves changes in protein concentrations relative to those of potential binding partners, where an imbalance in expression levels may potentially result in deleterious effects such as improper protein complex assembly. In order to provide balance in the system, duplicate genes may undergo various paths for divergence from its ancestor such as point mutations, gene fusions, truncations, deletions and gene conversions. Mutation of a duplicate gene can arise through multiple avenues that allow for divergence in a manner that includes alterations in gene expression patterns through promoter mutations, changes in ligand specificity, gain/loss of protein domains and insertion/deletions of short linear motifs (figure 5b). Many factors influence the number and divergence patterns for duplicate genes. For example, protein–protein interactions of the ancestral gene can influence the number of duplicates. The more highly connected and more centralized proteins found in humans are usually encoded by duplicated genes, which favour the functional diversification of paralogues [91]. This feature is unique to vertebrates following whole genome duplication (WGD), as gene dosage in these cases, is commonly controlled through tissue-selective expression and miRNA regulation. High connectivity may favour the functional diversification of paralogues through tissue specialization or other mechanisms [92,93]. A strong correlation \( r = 0.93 \) of the number of cell types with the number of SH2 domains suggests that duplication may have driven tissue-specific expression enabling specialized tissues and development programmes to evolve [94]. Dosage balance may act as an intermediate step to neo- and subfunctionalization, prolonging the retention of the duplicates before one of the other mechanisms determines the ultimate fate of the duplicates. This highlights the important role of functional divergence of duplicate genes in optimizing pTyr networks for specialized tissues and creating novel signalling pathways.

Examination of the 38 conserved SH2 families reveals 33 families with at least one gene duplicate in the human genome. Within these families, there are notable examples of some of the mechanisms driving divergence between duplicate copies. Duplication can lead to loss or gain of domains, often at the ends (N-terminus or C-terminus) of the protein [95]. In addition, duplicate genes can diverge in sequence to acquire novel binding motifs either within the SH2 domain or elsewhere within the protein (discussed in detail in §7). Five generalized classes have been described for the divergence of SH2 families as a result of duplication events combined with domain loss or gain [8]. Gene duplication can simply result in multiple members of a family, though the size of the family can vary substantially with the SRC family having eight functional genes in humans. The heterogeneity in family size likely relates to heterogeneity in both the gain and loss of duplicate genes between families, though it is not fully understood what evolutionary forces determine the pattern of gene retention. In some cases, a novel domain may be inserted or lost such as a SAM domain in SLP76 or the UBA domain in Cbl-C, respectively. Some families arise from neofunctionalization, where the gene family undergoes a change such as a loss of catalytic domain, as in the case of SH2D1 and SLAP, followed by an additional level of duplication for subfunctionalization. Lastly, several SH2 families such as RASA1 and SH3BP2 remain single copy, suggesting that any duplications that arose did not undergo fixation or that the genes were inactivated through non-functionalization. These classes, while general in definition, encompass the mechanism for divergence by gene duplication.

7. PHOSPHOTYROSINE NETWORKS EVOLVE THROUGH SH2-DOMAIN-MEDIATED INTERACTIONS

The pTyr signalling networks are important contributors to the evolution of organismal complexity in the unikont branch of Eukaryota, and particularly in Metazoa. Genome-wide loss of tyrosines in proteins across metazoans correlates with the expansion and utilization of pTyr as a novel PTM that carries significant evolutionary forces have driven the assembly of novel regulatory pathways and networks in pTyr signalling over the course of metazoan evolution. The pTyr protein interaction network involves the physical interaction between a pTyr ligand and the recognition by domains such as SH2 and PTB. Existing interactions that provide a selective advantage must either coevolve or remain conserved through evolution. Novel interactions may evolve rapidly by integrating pTyr motifs recognized by SH2 domains into scaffolds and other SH2 domain proteins. The identification of both conserved and rapidly evolving phosphorylation sites may have important implications for understanding human diseases [96]. Understanding of the mechanisms driving the elaboration of pTyr networks by modification of SH2 domain specificity and the integration of pTyr signalling into diverse signalling pathways also have implications for synthetic biology approaches such as cellular rewiring and creating customized
Within the SH2 domain, an invariant arginine residue within the second β-strand is necessary for coordinating the negatively charged phosphate of pTyr. Specificity is then determined through the interactions with pockets on the surface of the SH2 domain that recognize the residues primarily C-terminus to the pTyr, but in various cases extending from the −2 position up to the +5 position in respect to the pTyr [98–100]. Loop regions of the SH2 fold play additional roles in ligand selectivity through their ability to block certain contact regions and limit ligand access to the binding surface [98]. Selectivity is not limited to direct contact of residues that impart binding energy, as longer range charge repulsion, steric clashes and entropic effects also play a role in selectivity. Interestingly, SH2 domains explore only a very limited sequence space in their potential ligands. Two large classes of SH2 domains exhibit a strong preference for either an asparagine residue at +2 (e.g. pY-X-N-X) or a proline or aliphatic residue at +3 (e.g. pY-X-X-P/L). Despite this limited apparent selectivity for permissive residues, necessary to impart the minimal energies required for binding, SH2 domains exhibit remarkable selectivity for physiological peptide ligands. The basis for this is the extensive use of non-permissive residues that distinguish the binding of SH2 domains that otherwise have very similar primary motifs for binding [101]. In this way, the Crk and Brk SH2 domains are able to distinguish ligands using non-permissive residues and sequence context [101]. In evolutionary terms, this implies remarkable flexibility in adaptive mechanisms. A newly arising SH2 domain may preserve the essential binding motif of its immediate ancestral parologue, with adaptation allowing fine-tuning of ligands without major disruption to the primary binding surface. The use of non-permissive residues is inherently more flexible than permissive residues that require direct contacts and constrained geometry of binding pockets. Non-permissive residues can be accommodated in various ways, ranging from steric clashes within or adjacent to the binding groove, longer range charge-repulsion, or even entropic penalties associated with altering the conformation of the peptide upon binding [101].

The fidelity of pTyr signalling is in large part determined by the ability of a given SH2 domain to selectively recognize a specific set of pTyr ligands. This involves the interplay between the PTK/PTP, the SH2 domain and the pTyr ligand (figure 6a). In many cases, this means that over evolutionary distances, SH2 domains display conservation of ligand specificity, thereby maintaining core-binding partners whose ligand sequences also do not diverge significantly. Novel binding partners may still arise by insertion of pTyr motifs or mutations of sequences surrounding the pTyr to permit or abolish binding by specific SH2 domains. A complementary mechanism is for specificity to evolve through changes in the binding surface of an SH2 domain such that it recognizes different ligands than its ancestral relatives, allowing for novel interactions and re-wiring of the network. The obvious issue with altering the primary binding interface is that existing interactions will be lost, potentially with deleterious consequences for organismal fitness. Thus, large changes within the primary binding pocket are uncommon and the peptide binding groove is generally highly conserved among orthologous SH2 domains across various species [8]. Duplication events provide the possibility for more significant alterations, particularly if deleterious dosage effects provide the selection pressure to adopt alternative binding partners. More subtle alterations to the binding partner repertoire may be achieved with changes outside the primary contact region that impart restrictions on the binding partners, such as prohibiting a specific residue at a specific position. The refinement of non-permissive residues that can arise from changes outside the primary pocket can extend to distinguishing physico-chemically related amino acid residues. Thus, the Brk SH2 domain treats a glutamic acid at +1 as a permissive residue, whereas an aspartic acid is non-permissive, the Crk SH2 binds to ligands with a +3 proline or leucine residues but not to ligands containing the closely related isoleucine or valine residues, and Grb2 SH2 binds to ligands containing pY-E-N-E but not pY-D-N-D, etc. [101] (figure 6b). Such subtlety in distinguishing binding by related SH2 domains is consistent with the notion that specificity of SH2 domains coevolves with its pool of available ligands to fine-tune specificity over time as the size and complexity of pTyr signalling networks increases. For SH2 families, such as GRB2 (consisting of Grb2, Gads and Grap), NCK (Nck1 and Nck2) and CRK (Crk and CrkL), the specificity pocket appears highly conserved from fly to human [8] and these therefore share a common core set of binding peptides from fly to humans [101] (Nash laboratory 2012, unpublished data). In addition, insertion of novel binding sites within SH2 domain proteins is prevalent, particularly between arthropod and mammal, in a manner that serves to create more highly connected and potentially robust pTyr interaction networks. The Grb2 family encodes two SH3 domains flanking the central SH2 domain (table 1 and figure 6c). Examination of the conservation of the specificity pockets of the two SH3 domains reveals high conservation of the N-terminal SH3 domain pocket but much lower conservation of the C-terminal SH3 domain ligand binding surface, suggesting that this domain may have evolved to recognize novel binding partners that distinguish its function between arthropod and mammals. Indeed, the C-SH3 of the GRB2 family recognizes a very specific and atypical PxxxxRxxKxP motif in which the invariant and essential residues are the RxxK [102–104]. This differs substantially from the canonical PxxP motif recognized by most SH3 domains, and suggests that mutations in the C-SH3 mutated to accommodate a novel binding modality that became fixed owing to selective interactions with a handful of proteins harbouring the extended RxxK motif. While these examples demonstrate both conservation and adaptations in specificity, certain SH2 domains appear to have lost the ability to recognize pTyr ligands at all. For example, individual members of the RIN (Rin2) and JAK (Tyk2) family harbour a mutation of the essential βB-Arg to a His residue, presumably abolishing pTyr ligand binding [3].
Figure 6. Evolution of protein domain specificity. (a) The specificity of a protein interaction domain, such as an SH2 domain, can either change or remain conserved over time. Conservation of peptide ligand specificity pocket allows interactions to be conserved throughout evolution. Insertion of linear motifs and evolving novel binding sites can allow for an SH2 domain to recognize multiple binding partners. Specificity can also evolve for a domain, yet this can result in acquiring a completely new set of binding partners while the ancient partners are lost. Lastly, a domain can coevolve over time to maintain ancient binding partners but also evolve novel binding partners by evolving its peptide-binding pocket. A protein domain can become inactivated through mutations of its binding pocket such that it no longer can recognize any of its original ligands. (b) SH2 domains use sequence context and non-permissive residues to discriminate pTyr peptide ligands. This underlies a wider channel for information flow between the ligand and the SH2 domain. For example, the SH2 domains of Brk and Crk can each engage a subset of peptides containing a Pro or Leu at +3 (three residues C-terminus to pTyr; left panel): Brk and Crk distinguish between peptides containing a +3 Pro through distinct recognition of permissive and non-permissive residues at +1 and +4. Brk SH2-domain binding is favoured with permissive factors for Brk (+1 Glu and +4 Glu), which are also non-permissive factors for Crk favour Brk binding (middle panel). A +1 Asp and +4 Arg favours Crk binding over Brk (right panel). This is suggestive of coevolution between SH2 domains and peptide ligands to maximize specific recognition events using available peptide sequence information. (c) The adaptor Grb2 protein contains two SH3 domains (N-SH3, C-SH3) and a central SH2 domain. The sequence alignment of the GRB2 family from C. elegans to H. sapiens was mapped onto the structure of each individual domain. The percentage sequence conservation of the three domains is represented as a heat map on the surface of the solved ligand-bound structures (see legend for percentage of conservation). Structures used to highlight conservation are PDB: 1BMB; PDB: 2W0Z; PDB: 1GBR.
addition to evolving SH2 domain specificity through its ligand-binding pocket, domains may evolve secondary binding sites with their ligands to improve affinity, specificity or alter the biophysical nature of the interaction (e.g. by reducing off-rate to enhance downstream signalling). For example, the N-SH2 domain of phospholipase Cγ uses secondary contacts through its BC and DE loops for interaction with the C-lobe of the fibroblast growth factor receptor (FGFR) catalytic domain structure, which are important for binding to pY766 in FGFR1 [105]. SH2 domains may also exert allosteric regulation over other aspects of their host protein. For example, intramolecular contacts between the SH2 domain and tyrosine kinase domain directly regulate the enzymatic substrate activity for SH2-domain-containing kinases such as Fes and Abl [106]. Thus, SH2 domain proteins can evolve in multiple fashions to achieve selectivity in pTyr-mediated signalling.

Following a WGD event, many duplicate copies of genes are lost, whereas a fraction of the duplicates survive. Immediately following the WGD event, regulatory networks may be rapidly rewired to integrate the newly duplicate genes and prevent chaotic loss of control over cellular processes. The resulting rapid evolution and functional divergence can occur, in part, at the level of transcription as mentioned earlier [92]. However, adaptation may also occur at the level of post-translational regulation. Sites of PTMs are typically short motifs that commonly occur within intrinsically disordered protein regions that are inherently under less selective pressure to maintain sequence and thus are capable of rapid evolution. Changes in PTM sites thus represents one mechanism of rapidly effecting necessary rewiring.

A study of phosphosites across paralogous yeast genomes following WGD identified duplicate genes as having on average more phosphosites than genes that were not duplicated [107]. This post-translational level of regulation might be a means of rapidly rewiring duplicate copies, allowing these genes to adapt and to be retained.

SH2 proteins typically contain an assortment of other modular domains and short linear motifs that prescribe a range of additional interactions. The product of a gene duplication may give rise to a new gene that preserves
certain functions of the parent while acquiring specialized functions to coordinate novel interactions [108]. For example, the Socs proteins have evolved to fine-tune their primary binding interface in order to recognize specific sets of targets. Thus, Socs2 binds specifically to the growth hormone receptor, whereas Socs4 recognizes phosphorylated EGFR [109]. Both Crk and Sh2d1a each evolved as duplications of existing family members and then acquired new interactions through insertion of extended loop regions within the SH2 domain that act to promote the recruitment of SH3 domains via a non-pTyr-dependent, secondary binding surface within the SH2 domain. The gene family CRK encodes two paralogues, Crk and CrkL, and in mammals the Crk SH2 domain evolved an extended proline rich loop between the βD and βE strands of the SH2 domain that specifies an interaction with the SH3 domain of the CTK c-Abl [78,110] (figure 7). The insertion of a splice site within the βD–βE loop in vertebrate Crk but absent in CrkL allowed for evolutionary divergence through extension of this loop [8]. Splice site slippage is one way that such segment fragments may be inserted [111], allowing for rapid divergence between paralogues [8]. Evolving new binding sites is both a mechanism for diversifying duplicate genes as well as enabling the development of more complex signalling networks within specific tissues. Collectively, these mechanisms suggest that gene duplication may have served in part to promote SH2 domain proteins to coevolve and diverge within tissue types to create specialized cellular signalling networks such as those involved in signalling in lymphocytes.

8. CONCLUSIONS
The evolution of pTyr signalling involves the complex interplay of signalling components for control and activation of functional downstream signalling events through phosphorylation of tyrosine residues by PTKs and the recognition of specific pTyr sites by SH2 or other pTyr-binding domains. The site of phosphorylation and the sequence context surrounding the pTyr is under evolutionary pressure both to optimize kinase/phosphatase specificity as well as binding specificity for SH2 or other domains. Therefore, SH2 domains must coevolve with pTyr sites to optimize the appropriate level of specificity in signalling. Clearly, PTPs also play a role in the evolution of this system and present a further opportunity for study. There are still many unanswered questions, such as what are the primordial functions of SH2 domain proteins in their earliest forms? What were the driving forces for diversifying and expanding these classes of proteins at specific times? And what role did pTyr signalling play in the emergence and elaboration of multicellularity in the metazoan lineage? It is likely that important insights into the role of SH2 domains will come from future dissection of biological networks and pathways through genetics in lower organisms. Additional sequencing of genomes representing early branches of the unikont lineage and early steps in the development of metazoan multicellularity as well as biochemical analysis of pTyr signalling pathways in these organisms may provide deeper insight into the origins and evolution of pTyr signalling networks.

Addendum: During the publication process, two relevant articles were published that provide further insight into phosphotyrosine evolution that would be remiss to ignore. Jeffrey G. Williams group identified Pyk2 as the kinase responsible for phosphorylation and activation of STXIC in Dictyostelium (Proc. Nat. Acad. Sci. PMID: 22699506). Patrick Cramer and Dirk Eick's groups showed the tandem SH2 domains of SPT6 bound well to pTyr1, pTyr1/pSer2, pTyr1/pSer5 but not to unphosphorylated CTD. This indicates that Tyr1 phosphorylation stimulates CTD binding of the bona fide elongation factor SPT6 (Science, PMID: 22745433).

We thank Tony Pawson, Chris Tan and the members of the Pawson laboratory for helpful discussions. B.A.L is supported by a Canadian Institutes of Health Postdoctoral Fellowship. P.D.N. is supported by funds from the National Science Foundation grant (no. MCB-0819125), the University of Chicago Cancer Research Center and the Cancer Research Foundation.

REFERENCES


Phil. Trans. R. Soc. B (2012)


mammalian Grb2 couple the EGF receptor to the Ras activator mSos1. *Nature* **365**, 83–85. (doi:10.1038/363083a0)


900–911. (doi:10.1038/nri1226)


978. (doi:10.1073/pnas.0710316105)


102 Lewitzky, M. et al. 2001 The C-terminal SH3 domain of the adapter protein Grb2 binds with high affinity to sequences in Gab1 and SLP-76 which lack the SH3-typical P-x-x-P core motif. *Oncogene* **20**, 1052–1062. (doi:10.1038/sj.onc.1204202)


