The JAK-STAT pathway represents a finely tuned orchestra capable of rapidly facilitating an exquisite symphony of responses from a complex array of extracellular signals. This review explores the evolution of the JAK-STAT pathway: the origins of the individual domains from which it is constructed, the formation of individual components from these basic building blocks, the assembly of the components into a functional pathway, and the subsequent reiteration of this basic template to fulfill a variety of roles downstream of cytokine receptors.

Introduction

The ability to respond to external signals is essential for life. At the level of the whole organism these signals are principally captured by peripheral sensory organs, which transmit them via the nervous system to the brain to mediate an appropriate response. At the cellular level, these signals are generally received by a myriad of cell-surface receptors that utilize a range of intracellular signaling pathways to communicate with the nucleus. This rapidly impacts on target gene transcription and mediates the desired cellular outcome, such as proliferation, differentiation, activation, and survival. The Janus kinase–signal transducer and activator of transcription (JAK-STAT) pathway represents a major mechanism used to transmit signals from extracellular receptors to the nucleus. The canonical pathway consists of JAK and STAT proteins possessing an S-type immunoglobulin fold seen in members of the "p53 clan"—that includes p53 and RUNT—which mediates sequence-specific binding, while the coiled-coil domain is involved in protein–protein interactions. Finally, the transactivation domain (TAD)—of which there are several varieties—is important for co-factor recruitment and transcriptional responses.

These domains are assembled to form the individual components of the JAK-STAT pathway (Fig. 1B):

JAK-STAT Pathway Components

To understand the evolution of the various components of the canonical JAK-STAT pathway, it is necessary to be familiar with their constituent protein domains, and how these modules are assembled into individual components (Fig. 1).

At the core of the JAK-STAT pathway is a binary code, based on specific tyrosine residues that are either unphosphorylated (off) or phosphorylated (on). Three of the protein domains utilized in the JAK-STAT pathway are specifically involved in the writing, reading and erasing of this code: the protein tyrosine kinase (PTK), Src-homology 2 (SH2) and protein tyrosine phosphatase (PTP) domains, respectively (Fig. 1A). The PTK domain consists of ~250–300 residues, including an ATP-binding site and a catalytic region, and is able to phosphorylate tyrosine residues with a specificity determined by access to substrate and sequence context, thereby writing the code. The SH2 domain is a ~100 residue motif that binds to phosphotyrosine residues in a manner dependent on the specific residues around the tyrosine especially those downstream, providing a mechanism for reading the code. Finally, the PTP domain consists of ~280 residues, with a signature CX_R motif, and is capable of dephosphorylating phosphotyrosine residues with which it is able to effectively interact, thereby erasing the code.

JAK-STAT signaling components also consist of several other domains that provide accessory functions. The FERM domain, named after the founding proteins identified with this domain (band 4.1, ezrin, radixin and moesin), consists of three subdomains, F1–F3, that are structurally similar to ubiquitin, acyl-CoA binding and plecstrin homology-phosphotyrosine binding domains, respectively. The FERM domain is able to mediate protein–protein interactions, including adaptor and scaffolding interactions with membrane bound proteins. FERM domain containing proteins can be divided into three broad groups: (1) talins and kinesins, (2) ERMS, GEF, kinases and phosphatases and (3) myosins and KIRT. The SOCS box is a ~40 residue motif that mediates interactions with proteasomal degradation pathway components—particularly ubiquitin ligases—thereby regulating protein half-life. The DNA-binding domain present in STAT proteins possesses an S-type immunoglobulin fold seen in members of the "p53 clan"—that includes p53 and RUNT—which mediates sequence-specific binding, while the coiled-coil domain is involved in protein–protein interactions. Finally, the transactivation domain (TAD)—of which there are several varieties—is important for co-factor recruitment and transcriptional responses.

Keywords: JAK, STAT, SOCS, SHP, evolution

Citation: Liongue C, Ward AC. Evolution of the JAK-STAT pathway. JAK-STAT 2013; 2:e22756; http://dx.doi.org/10.4161/jkst.22756

Submitted: 10/19/12; Accepted: 11/02/12

*Correspondence to: Alister C. Ward; Email: award@deakin.edu.au

School of Medicine and Strategic Research Centre in Molecular & Medical Research; Deakin University; Geelong, VIC Australia
a catalytic PTP domain and a divergent C-terminal region that contains several tyrosine residues that serve as docking sites for other signaling proteins when phosphorylated.11,23

**SOCSs:** Consist of a divergent N-terminal region of largely unknown function, a central SH2 domain responsible for binding to cytokine receptors, JAKs and other substrates and a C-terminal SOCS box domain that facilitates protein degradation of the cytokine receptor signaling molecules.24

Specific SOCS members also possess additional motifs, such as the kinase inhibitory region (KIR) of SOCS1 and SOCS3, and the N-terminal conserved region (NTCR) of SOCS4 and SOCS5.25,26

The **JAK-STAT Pathway**

The individual components interact functionally in the canonical JAK-STAT pathway (Fig. 2). This pathway is activated by cytokine receptors in response to extracellular ligand binding, which leads to conformational changes in the receptor that

---

**Figure 1. Architecture of JAK-STAT pathway components.** (A) The constituent domains of JAK-STAT pathway components and their functionality. The PTK, SH2 and PTP domains contribute to the write, read and erase functionalities of the JAK-STAT pathway, as indicated, while the others have accessory roles. (B) The JAK-STAT pathway components and their domain assemblage.
activate the intracellularly associated JAKs. These phosphorylate the receptor complex to recruit STAT proteins, which are in turn phosphorylated, allowing dimerization and subsequent translocation to the nucleus to influence the transcription of relevant genes. The activated JAK-STAT pathway is then extinguished by negative regulators recruited to the receptor signaling complex, including latent SHPs or induced SOCSs. Additionally, the individual components are also able to signal via alternate “non-canonical” mechanisms.27,28

Building the Building Blocks

The first obligatory step in the evolution of the JAK-STAT pathway was the generation of each of the individual domains that make up the various components of the pathway. Only 136 protein domains are metazoan-specific, with the majority of protein domains (> 400) emerging early in eukaryote evolution,29 which is also the case for the constituent domains of the JAK-STAT pathway (Fig. 3).

PTK domain. There is genomic evidence that PTK domains emerged in early eukaryotes or indeed prior to eukaryotes, as a modification of pre-existing kinase domains with alternate amino acid specificity.16,30-32 PTK domains are certainly present in a broad range of extant non-metazoans, such as amoebozoa (such as the slime mold, Dictyostelium discoideum), basal holozoa (such as the filasterea, Capsaspora owczarzaki) and choanoflagellates (e.g., Monosiga brevicollis), and have become extremely widespread in the metazoan lineage due to the myriad of pathways that exploit tyrosine phosphorylation.33

SH2 domain. This domain, present in all JAK-STAT pathway components, can be traced to the origins of eukaryotes. Proto-SH2 domains are found in present-day plants (e.g., Arabidopsis thaliana) and slime molds, with the number of SH2 domains containing proteins expanding to > 100 in higher vertebrates.8,16,34 Interestingly, the SH2 domains of the JAK and STAT families are closely related when compared with those of other SH2 domain-containing proteins.7,35

PTP domain. The PTP domain is present in prokaryotes, bacteria and archaea, and also basal eukaryotes such as malaria (e.g., Plasmodium berghei), and indeed may have also been captured by several viruses.16 Like the PTK domain, this domain seems to have evolved by modification of a pre-existing enzyme to one with altered specificity for phosphotyrosines. Curiously, there are relatively few examples of PTP domains in extant representatives of plants and slime molds.36-38 However, the PTP domains have expanded considerably since the divergence of holozoa,39 again reflecting the increasing importance of phosphotyrosine-mediated signaling along this lineage.

Figure 2. The canonical JAK-STAT pathway. (A) Ligand binding to a cytokine receptor results in dimerization and conformational changes that causes activation of JAKs associated with the Box 1 and Box 2 domains of the receptor intracellular domain. (B) The activated JAK is able to phosphorylate tyrosines on the receptor complex to create docking sites for signaling proteins, such as STATs. These are in turn phosphorylated, and then dimerize and translocate to the nucleus, where they bind to specific sequences in the promoters of responsive genes. (C) The pathway is extinguished by pre-existing SHPs, which serve to dephosphorylate tyrosine residues, or induced SOCSs that can act via several mechanisms, including JAK inhibition, steric interference at STAT docking sites, or mediating degradation of receptor signaling components.
and early holozoans. Finally, proto-TAD domains probably arose in the bilateria.16

Component Assembly

The next important step in JAK-STAT pathway evolution was the assembly of these specific domains to generate each individual component. The assembly of domains into new protein architectures involves genomic processes such as exon-shuffling, exonization, intronization and duplication,42-44 while sequence divergence can modify these domains.29,44 These processes, rather than de novo generation of novel domains represents the major mechanism for the domain variation of proteins in eukaryotes, which is affirmed in the analysis of individual JAK-STAT pathway components.

JAKs.

Each of the constituent domains of the JAKs, with the exception of the pseudo-kinase domain, was present throughout eukaryote evolution. So, how did the unique JAK architecture arise? The FERM domain of JAK is most closely related to that of focal adhesion kinase (FAK), which consists of several components including the FERM domain. The FERM domain is found in present-day plants and basal eukaryotes,14,16,40 indicating a long evolutionary history. Indeed, there are six FERM domain containing proteins in present-day slime molds, although the ERMs, GEF, kinases and phosphatase group is absent,14 with a total of 22 FERM domain containing proteins in extant insects, such as the fruit fly (Drosophila melanogaster).13

SOCS box. Current genomic evidence suggests that the SOCS box is holozoan specific, with a SOCS box containing protein seen in present-day choanoflagellates (e.g., Salpingoeca rosetta).16 The number of SOCS box containing proteins has subsequently expanded considerably within metazoans.14

Other STAT-specific domains. The individual domains of STAT have arisen at different times during evolution. The p53 clan-type DNA binding domain has been present since early in eukaryote evolution,16 with such domains described in extant plants,41 early unikonta, such as amoebozoa, and holozoans, including choanoflagellates and ichthyosporea.16 Although the roots are not as ancient, the coiled-coil domain arose prior to the divergence of amoebozoa, being evident in present-day slime molds and early holozoans. Finally, proto-TAD domains probably arose in the bilateria.16
of an N-terminal FERM domain, SH2-binding domain (not actually a SH2 domain), PTK domain and a C-terminal focal adhesion targeting (FAT) domain, which is evident in basal holozoans, such as present-day filasterea (e.g., C. owczarzaki).

Interestingly, the JAK SH2 domain shows highest homology with the N-terminal SH2 domains of two other kinases, SYK and ZAP70, which possess dual SH2 domains and an adjacent PTK domain. The earliest SYK-like protein is exemplified by that in present-day sponge (e.g., Amphimedon queenslandica). It remains unclear whether the JAK architecture was generated by the addition of an SH2 domain to a FAK-like protein that subsequently lost the FAT domain, or the addition of the FERM domain to a SYK-like protein, or some fusion of a FAK and SYK protein, prior to the divergence of eumetazoa and porifera. Curiously, M. brevicollis contains a protein, UTK05mod, which consists of FERM, SH2 and dual kinase domains, that is reminiscent of the metazoan JAK proteins, although the kinase domain forms a clade with the SYK kinase domain and not that the JAK kinase domain. However, the similarity of the architecture between UTK05mod and JAK may be due to coincidence as there exist a wide variety of kinase domain containing architectures that are specific to M. brevicollis. The first bona fide JAK-like protein is represented in the basal metazoan porifera lineage, typified by the sequence BAA81718.3 found in the present-day sponge, Ephydatia fluviatilis, which consists of FERM and SH2 domains as well as a single PTK domain related to the JAK PTK, although lacking the pseudo-PTK domain. The canonical JAK domain architecture inclusive of a pseudo-kinase domain appears to be unique within bilateria, with this unique topology typified by present-day fruit fly Hopscotch. It is likely that that the pseudo-kinase domain a result of local duplication and divergence of the PTK domain within a JAK-like sequence, although there exists the possibility that it was by modification of the first PTK domain of a UTK05mod-related sequence.

STATs. The STAT family has more ancient roots, with a range of STAT-like proteins present in extant eukaryotes. The most divergent are the GRAS proteins of the plant kingdom (named after the founding members, GAI, RGA and SCR), which possess an SH2-like domain that is remarkably similar to the metazoan STAT SH2 domain and also a putative DNA binding domain. By the time of divergence of the amoebozoa, a family of proteins can be identified that resemble metazoan STATs, with coiled-coil, DNA-binding and SH2 domains, although lacking trans-activation domains and the N-terminal sequence, as exemplified by those in present-day slime molds. Extant members of the cnidaria, the sea anemone (Nematostella vectensis) and nematode (Caenorhabditis elegans), both contain at least one STAT with architectures similar to that seen in slime mold. Furthermore, the nematode STAT paralogs also contain a TAD, while the sea anemone STAT contains an N-terminal sequence, but not a TAD, although it possesses a long C-terminus containing multiple tyrosine and serine residues. The canonical STAT, replete with both TAD domain and N-terminal sequence, first appeared in the bilateria, as exemplified by present-day fruit fly Marelle (STAT92E).

SHPs. The constituent domains of the SHPs were also present throughout eukaryotes evolution, but there is no evidence of these domains coming together in the bikonta or amoebozoa. However, a hypothetical SHP-like protein, is found in present-day choanoflagellates (e.g., XP_001747983 in M. brevicollis), consisting of a single SH2 domain followed by a PTP domain, which suggests that this architecture had evolved by the time of holozoan. The canonical SHP was finalised early in metazoan evolution, with the extant nematode ortholog, ptp-2, and fruit fly Corkscrew displaying the unique dual N-terminal SH2 domain architecture. This most likely arose via local duplication of the adjacent SH2 domain in the precursor as the N-terminal and C-terminal SH2 domains of SHP are more similar to each other than the SH2 domains of other proteins.

SOCSs. Evolution of the SOCS proteins has followed a slightly more unique path. The most ancient SOCS box-containing protein architecture consisted of three N-terminal ankyrin repeats and a C-terminal SOCS box, similar to that of metazoan ankyrin repeat and SOCS box (ASB) proteins. The SOCS SH2 domain is most closely related to that of dual adaptor of phosphotyrosine and 3-phosphoinositides (DAPP1), which together are part of a much larger clade that includes the SHP SH2 domain. Presumably domain accretion to fuse a SH2 domain with a SOCS box generated the archetypical SOCS. In contrast to other components, this prototype underwent several gene duplication events that established three SOCS members by the divergence of the cnidaria. This is exemplified in extant cnidaria (e.g., N. vectensis), which possess three proteins with the signature SOCS protein architecture. These are related to each of the three distinct SOCS clades seen in higher vertebrates: SOCS1/2/3/CISH, SOCS4/5 and SOCS6/7.

When JAK Met STAT and SOCS and SHP: The Canonical JAK-STAT Pathway

At the time of the emergence of bilateria, archetypal members of each of the JAK-STAT pathway components had been generated, including a single JAK, STAT and SHP, and three SOCS proteins. These came together to form the canonical JAK-STAT signaling pathway prior to the divergence of protostomia, as exemplified by the present day fruit fly Hopscotch-Marelle-Corkscrew/Socs36E pathway downstream of Domeless receptors. Several of the key innovations that were required to bring the various components together were, first, the association of the prototypal JAK with a membrane-bound receptor that lacked intracellular tyrosine kinase activity, second, appropriate tyrosines with which to recruit STATs, SHPs and SOCSs, and third, the incorporation of a negative feedback loop via the SOCS family. These innovations provided the ability to both recruit and activate the writer, the “paper” on which to write and a vehicle to inducibly block the writer or reader and/or act as a “paper shredder” to extinguish the signal, thereby complementing the erasers. Surprisingly little was required to establish these innovations. A few mutations to generate a cluster of intracellular prolines would have been sufficient for a cytokine receptor to be able to interact with the FERM domain of a prototypal JAK. A similar process could generate
tyrosines in an appropriate sequence context, and a small number of mutations in the promoter region of a pre-existing SOCS gene would have been enough to generate an appropriate STAT binding site (TTCN$_{3-4}$ GAA) to enable their recruitment into an inducible feedback loop. Of course, each of the components would have necessarily participated in other processes prior to their recruitment into the canonical JAK-STAT pathway. Indeed, the legacy of this is evident in some of the non-canonical roles played by the pathway components in alternate signaling paradigms, such as downstream of receptor tyrosine kinases.

Making the Most of a Good Thing

Since the establishment of the canonical JAK-STAT pathway, it has subsequently been replicated multiple times to establish the diversity seen within extant vertebrates. However, the pathway itself has not been replicated per se, but rather the individual components have undergone duplication, creating an opportunity for the generation of new pathway combinations.

Duplication events, either local or global, have largely facilitated the expansion of gene families in vertebrates. At the global level, three whole genome duplications (WGDs) have molded vertebrate evolution: the so-called 1R and 2R, which occurred before the divergence of lobe-finned and ray-finned fishes, and 3R, which was limited to teleost fish (Fig. 3). While rapid loss of many duplicated genes occurred following WGD, some have been retained because of either division of function between the duplicates or the generation of new functions for one of the duplicates. We and others have shown that WGDs supplemented by local duplications have generated the diversity of JAK-STAT pathway components seen in higher vertebrates.

Following the establishment of the archetypal JAK (typified by fruit fly Hopscotch) prior to the divergence of protostomes and deuterostomes, there has been no increase in the number of JAK homologs until the emergence of vertebrates, with existing urochordates such as sea squirt (Ciona intestinalis) retaining a single JAK. The four JAKs in present-day tetrapods (e.g., Homo sapiens) — JAK1, JAK2, JAK3 and TYK2 — were generated from 1R and 2R via JAK1/TYK2 and JAK2/TYK3 intermediates, while 3R yielded the additional teleost-specific JAK2 paralogs, named jak2a and jak2b, such as in zebrafish (Danio rerio).

The archetypal STAT (typified by fruit fly Marelle) was similarly finalized by the divergence of protostomes and deuterostomes. However, local duplication appears to have generated distinct duplicates in chordates prior to divergence of the urochordates, as exemplified in present-day sea squirt. WGDs have mediated the on-going expansion of the family within vertebrates. SOCS1, SOCS2, SOCS3 and CISH were generated from the SOCS1/2/3/CISH precursor during 1R and 2R, via SOCS1/SOCS3 and SOCS2/CISH intermediates. Concurrent gene loss limited the expansion of the other genes to yield SOCS4 and SOCS5 from the SOCS4/5 precursor and no further expansion of SOCS6 and SOCS7. In contrast, the 3R event produced paralogs for SOCS3 (socs3a and socs3b), SOCS4 (socs4a and socs4b), SOCS5 (socs5a and socs5b) and CISH (cish.a and cish.b), as seen in the present-day teleost, D. rerio.

Loss of duplicated genes is attributed to the generally redundant functions of the duplicate. Therefore, appropriate selective pressure is required to maintain duplicates. We have hypothesized that the key selective force in the case of JAK-STAT pathway component retention has been the expansion of the upstream cytokine receptors. This occurred co-incidently with an increase in organism complexity, including the emergence of adaptive immunity, providing an opportunity for the duplicates to develop crucial, non-redundant functions that facilitated retention.

Conclusion

There have been four critical phases in the evolution of the complex canonical JAK-STAT pathway: (1) the generation of the individual domains with discrete functions, such as the writing, reading, and erasing of a phosphotyrosine code, or other accessory roles, most of which occurred early in eukaryotic evolution, (2) the arrangement of these domains into functional proteins, which happened gradually across many years, (3) the interaction of these proteins to form an active pathway, which occurred in bilateria, and (4) the subsequent proliferation of this pathway, which was largely driven by WGD during vertebrate evolution. This has collectively facilitated the generation of a complex JAK-STAT pathway able to orchestrate responses to appropriate extracellular signals to produce the coherent melody of transcriptional outcomes that underpins hematopoiesis, immunity, growth and other essential functions in higher vertebrates.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.
49. Dierking K, Polanowska J, Omi S, Engelmann I, Gut M, Lembo F, et al. Unusual regulation of a STAT protein by an SLC6 family transporter in C. elegans epidermal innate immunity. Cell Host Microbe 2011; 9:425-35; PMID:21575913; http://dx.doi.org/10.1016/j.chom.2011.04.011

50. Putnam NH, Srivastava M, Hellsten U, Dirks B, Chapman J, Salamov A, et al. Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 2007; 317:86-94; PMID:17615350; http://dx.doi.org/10.1126/science.1139158

51. Wang Y, Levy DE. C. elegans STAT: evolution of a regulatory switch. FASEB J 2006; 20:1641-52; PMID:16873887; http://dx.doi.org/10.1096/fj.06-6051com

52. Gurch MJ, Flint AJ, Keller J, Tunks NK, Hengartner MO. The Caenorhabditis elegans SH2 domain-containing protein tyrosine phosphatase PTP-2 participates in signal transduction during oogenesis and vulval development. Genes Dev 1998; 12:571-85; PMID:9472025; http://dx.doi.org/10.1101/gad.12.4.571

53. Liongue C, John LB, Ward A. Origins of adaptive immunity. Crit Rev Immunol 2011; 31:61-71; PMID:21395312; http://dx.doi.org/10.1615/CritRevImmunol.v31.i1.60

54. Shi S, Calhoun HC, Xia F, Li J, Le L, Li WX. JAK signaling globally counteracts heterochromatic gene silencing. Nat Genet 2006; 38:1071-6; PMID:16892059; http://dx.doi.org/10.1038/ng160

55. Shi S, Larson K, Guo D, Lim SJ, Dutta P, Yan SJ, et al. Drosophila STAT is required for directly maintaining HP1 localization and heterochromatin stability. Nat Cell Biol 2008; 10:489-96; PMID:18344984; http://dx.doi.org/10.1038/ncli1713

56. Usacheva A, Korenko S, Witte MM, Colamonici OR. Two distinct domains within the N-terminal region of Janus kinase 1 interact with cytokine receptors. J Immunol 2002; 169:1302-8; PMID:12133952.

57. Kouro T, Kikuchi Y, Kanazawa H, Hirokawa K, Harada N, Shiiba M, et al. Critical proline residues of the cytoplasmic domain of the IL-5 receptor alpha chain and its function in IL-5-mediated activation of JAK kinase and STAT5. Int Immunol 1996; 8:237-45; PMID:8671609; http://dx.doi.org/10.1093/iummun/8.2.237

58. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrates. PLoS Biol 2005; 3:e314; PMID:16128622; http://dx.doi.org/10.1371/journal.pbio.0030314

59. Jallon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, Mauzé E, et al. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature 2004; 431:946-57; PMID:15496914; http://dx.doi.org/10.1038/nature03025

60. Wolfe KH. Yesterday’s polyploids and the mystery of diploidization. Nat Rev Genet 2001; 2:333-41; PMID:11331899; http://dx.doi.org/10.1038/35072009

61. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, et al. The amphioxus genome and the evolution of the chordate karyotype. Nature 2008; 453:1064-71; PMID:18563158; http://dx.doi.org/10.1038/nature06967

62. Hino K, Satou Y, Yagi K, Satoh N. A genomewide survey of developmentally relevant genes in Ciona intestinalis. VI. Genes for Wnt, TGFbeta, Hedgehog and JAK/STAT signaling pathways. Dev Genes Evol 2003; 213:264-72; PMID:12739142; http://dx.doi.org/10.1007/s00427-003-0318-8

63. Hou XS, Melnick MB, Perrimon N. Marelle acts downstream of the Drosophila HOP/JAK kinase and encodes a protein similar to the mammalian STATs. Cell 1996; 84:411-9; PMID:8608595; http://dx.doi.org/10.1016/S0092-8674(00)81286-6