
Csaba Ortutay,* Beston F. Nore,† Mauno Vihinen,*‡ and C. I. Edvard Smith†

*Institute of Medical Technology, FI-33014 University of Tampere, Finland
†Department of Laboratory Medicine, Clinical Research Center, Karolinska Institutet, Karolinska University Hospital Huddinge, SE-141 86 Huddinge, Sweden
‡Tampere University Hospital, FI-33520 Tampere, Finland

I. Tyrosine kinases and the TEC family
   A. Identification and characteristics of TFKs
   B. Biological functions of TFKs
II. Aim of this Review
III. The function of individual domains in TFKs
   A. Function of the SH3–SH2–kinase domain complex
   B. Regulation of PH domain binding by phosphoinositide 3-kinase (PI3K)
   C. Regulation of TFKs through the PH domain by serine/threonine kinases
   D. Function of the TH domain
   E. Regulation of Btk through SH3BP5
   F. Mutations in X-linked agammaglobulinemia as a tool to study the function of residues
IV. The ancestry of TFKs
   A. Collecting the sequences
   B. Aligning the sequences
   C. Phylogenetic analysis
   D. The origin of the Btk-specific PH domain loop in amniotes
   E. N-terminal regions of insect TFKs
V. SH3BP5—A conserved negative regulator of TFKs
VI. The origin of phosphotyrosine signaling and the role of cytoplasmic tyrosine Kinases

Acknowledgments

References

ABSTRACT

It is generally considered mammals and birds have five Tec family kinases (TFKs): Btk, Bmx (also known as Etk), Itk, Tec, and Txk (also known as Rlk). Here, we discuss the domains and their functions and regulation in TFKs. Over the last few years, a large number of genomes from various phyla have been sequenced making it possible to study evolutionary relationships at the molecular and sequence level. Using bioinformatics tools, we for the first time demonstrate that a TFK ancestor exists in the unicellular choanoflagellate Monosiga brevicollis, which is the closest known relative to metazoans with a sequenced genome. The analysis of the genomes for sponges, insects, hagfish, and frogs suggests that these species encode a single TFK. The insect form has a divergent and unique N-terminal region. Duplications generating the five members took place prior to the emergence of vertebrates. Fishes have two or three forms and the platypus, Ornithorhynchus anatinus, has four (lacks Txk). Thus, not all mammals have all five TFKs. The single identified TFK in frogs is an ortholog of Tec. Bmx seems to be unique to mammals and birds. SH3BP5 is a negative regulator of Btk. It is conserved in choanoflagellates and interestingly exists also in nematodes, which do not express TFKs, suggesting a broader function in addition to Btk regulation. The related SH3BP5-like protein is not found in Nematodes.

2008, Elsevier Inc.

I. TYROSINE KINASES AND THE TEC FAMILY

A. Identification and characteristics of TFKs

In mammals, Tec family kinases (TFKs), Btk, Bmx (Etk), Itk, Tec, and Txk (Rlk), form the second largest family of cytoplasmic protein–tyrosine kinases (PTKs), the largest being related to Src, harboring nine Src family kinases (SFKs) (Caenepeel et al., 2004; Quintaje and Orchard, 2008). We have used the official Human Genome Nomenclature Committee (HGNC) (http://www.genenames.org/) gene and protein names throughout the text. Human TFKs are in upper case and gene names in italics. The 2008 annotation lists 480 classical and 24 nonclassical protein kinases in man, out of which 90 are PTKs, while mice have
93 tyrosine kinases (Quintaje and Orchard, 2008). The protein kinases constitute one of the largest mammalian gene families comprising about 2% of all genes or about 10% of signaling functions coding genes.

With the exception of Txk, TFKs are characterized by an N-terminal pleckstrin homology (PH) domain, followed by a Tec homology (TH), Src homology (SH) -3, -2, and -1 (catalytic) domains (Smith et al., 1994b; Vihinen et al., 1994a). Txk instead has a cysteine-rich string, which, like PH domains, is required for temporary membrane attachment. PH domains in TFKs are also involved in binding to heterotrimeric G proteins and protein serine/threonine kinases (PSKs).

The TH domain consists of an N-terminal Zn\(^{2+}\)-binding Btk motif (Hyvönen and Saraste, 1997; Vihinen et al., 1994a, 1997a) and one or two proline-rich motifs (Smith et al., 2001; Vihinen et al., 1994a, 1997a). Bmx lacks the typical proline-rich region (PRR) of the TH domain and also has an altered SH3 domain (Smith et al., 2001; Tamagnone et al., 1994). The PRRs participate in inter- and intramolecular SH3 domain binding. The SH2 and SH3 domains are docking modules, which bind to polyproline helices and phosphotyrosines, respectively. The kinase domain is the only catalytic entity in TFKs.

The mammalian TFKs were cloned during 1990–1995 (Haire et al., 1994; Heyeck and Berg, 1993; Hu et al., 1995; Mano et al., 1990; Robinson et al., 1996; Siliciano et al., 1992; Tamagnone et al., 1994; Tsukada et al., 1993; Vetrie et al., 1993; Yamada et al., 1993) and immediately received wide interest, especially owing to the fact that BTK mutations cause an X-linked form of B lymphocyte deficiency (X-linked agammaglobulinemia, XLA) in man (Lindvall et al., 2005; Tsukada et al., 1993; Väliaho et al., 2006; Vetrie et al., 1993; Vihinen et al., 1995a) and X-linked immunodeficiency (XID) in mice (Rawlings et al., 1993; Thomas et al., 1993). Btk is expressed in lymphoid cells but absent from T cells and plasma cells (Smith et al., 1994a).

Disease-related mutations affecting the corresponding enzyme in T lymphocytes, ITK, has to date not been reported, but this gene resides on an autosome making any loss-of-function phenotype considerably less common than that observed for BTK. However, Itk plays an essential role in T-lymphocyte development as shown by knocking out the gene in mice (Liao and Littman, 1995). The gene was initially identified using degenerate primers to amplify cDNA from an IL-2-dependent mouse T-cell line (Siliciano et al., 1992) or from neonatal mouse thymus (Heyeck and Berg, 1993).

Tec was cloned from a hepatic carcinoma (Mano et al., 1990), but was later found to be expressed in several tissues, including B lymphocytes (Mano et al., 1993). The phenotype of mice with Btk deficiency is much milder than the one seen in humans. While Tec is also expressed in human B lymphocytes, in mice the generation of double knockouts for Btk and Tec causes an XLA-like...
phenotype, whereas Tec single-knockout mice do not have an overt phenotype (Ellmeier et al., 2000). An osteoclast defect has been reported both in isolated Btk deficiency (Lee et al., 2008) and in combined Btk/Tec deficiency (Shinohara et al., 2008). More subtle effects of Tec have been recognized in platelets (Atkinson et al., 2003; Crosby and Poole, 2002; Oda et al., 2000), erythroid (Schmidt et al., 2004a; van Dijk et al., 2000), and phagocytic cells (Jongstra-Bilen et al., 2008; Melcher et al., 2008). In these cells, Tec is important for signaling through various receptors (Atkinson et al., 2003; Crosby and Poole, 2002; Oda et al., 2000; Schmidt et al., 2004b; van Dijk et al., 2000), and lack of Tec resulted in increased levels of caspases (Melcher et al., 2008). Furthermore, Tec showed a unique late effect in the phagocytic process (Jongstra-Bilen et al., 2008).

Bmx was originally identified from a bone marrow-derived cDNA library (Tamagnone et al., 1994) and was later found to be expressed mainly in endothelial cells as well as in prostate tumors (Ekman et al., 1997; Robinson et al., 1996). Loss-of-function mutations in humans have not been detected, which is somewhat unexpected owing to its X-chromosomal location and the viability of the mouse knockout. Such mice are characterized by defects in arteriogenesis and angiogenesis (He et al., 2006; Rajantie et al., 2001).

Txk was first identified from human peripheral blood and murine thymus cDNA libraries, respectively (Haire et al., 1994; Hu et al., 1995), and was later found to be mainly involved in T-lymphocyte development, since Itk\(^{−/−}\) and Txk\(^{−/−}\) mice have a more severe phenotype as compared to an Itk\(^{−/−}\) single defect (Broussard et al., 2006; Gomez-Rodriguez et al., 2007; Schaeffer et al., 1999). Recently a role for Txk has also been described in NKT-cell development (Felices and Berg, 2008).

### B. Biological functions of TFKs

The TFKs have been implicated as pivotal components of signaling pathways downstream of extracellular receptor stimuli, such as lymphocyte antigen receptors (Felices et al., 2007; Lindvall et al., 2005; Schmidt et al., 2004a,b). A functional defect of Itk and Btk kinases affect both innate and adaptive immunity in T cells (Berg et al., 2005) and B cells (Brunner et al., 2005; Hasan et al., 2008; Mansell et al., 2006), respectively. Activation of TFKs is a two-step event, which requires phosphorylation by a SFK member, and translocation to the plasma membrane, mediated by the PH domain (Lewis et al., 2001; Nore et al., 2000; Varnai et al., 2005). It is not known which of these comes first, but presumably membrane translocation is the initial event, since Btk mutants
lacking membrane-binding activity are not phosphorylated. Moreover, SFKs are known to mainly reside in the membrane (Ingle, 2008). In the case of Txk, which lacks PH domain, N-terminal palmitoylated cysteine-string motif is responsible for membrane colocalization.

Btk seems to have a dual role in apoptosis, under certain conditions being protective, while in other cellular contexts it instead induces apoptosis (Islam and Smith, 2000; Uckun, 1998). Bmx was reported to protect cells from apoptosis (Xue et al., 1999). The antiapoptotic role of TFKs may involve AP-1 signaling (Altman et al., 2004). Another potentially antiapoptotic pathway is through NF-κB, which is downstream of Btk (Bajpai et al., 2000; Petro et al., 2000). Recently was shown that NF-κB acts on the Btk promoter region by feedback activation (Yu et al., 2008).

One common phenomenon in TFK-mediated activation is that these kinases modulate actin polymerization and dynamics, which play a central role in cytoskeleton processes including cell division, motility, cell shape, and chemotaxis (Finkelstein and Schwartzberg, 2004; Gomez-Rodriguez et al., 2007; Nore et al., 2000). Btk29A in *Drosophila* is essential for head involution during embryogenesis and also for ring canal growth (Guarnieri et al., 1998; Roulier et al., 1998). Recent studies show that Btk29A controls actin remodeling (Chandrasekaran and Beckendorf, 2005) and microfilament contraction during embryonic cellularization (Thomas and Wieschaus, 2004). It is also of interest to note that the phenotype of flies lacking only the full-length isoform of Btk29A is not lethal (Baba et al., 1999). Instead, the development of male genitals and longevity are affected. The phenotype of TFK-deficient, more primitive species, Porifera (sponge) and choanoflagellate, is not known.

### II. AIM OF THIS REVIEW

A few reports have discussed the evolution of TFKs (Baba et al., 1999; Cetkovic et al., 2004; Guarneri et al., 1998; Haire et al., 1997, 1998; Nars and Vihinen, 2001; Roulier et al., 1998; Smith et al., 2001), but there are no recent publications related to this topic. Owing to the many genomes characterized over the last years, it seemed timely to compile and analyze the existing data.

SH3BP5 (also called Sab and in flies Parcas) has been identified as a negative regulator of at least Btk in both mammals and *Drosophila melanogaster* (Hamada et al., 2005; Matsushita et al., 1998; Sinka et al., 2002). For this reason the phylogeny of this regulator was also investigated. The emergence and functions of TFKs as well as SH3BP5 are discussed from an evolutionary point of view.
III. THE FUNCTION OF INDIVIDUAL DOMAINS IN TFKs

A. Function of the SH3–SH2–kinase domain complex

Apart from TFKs, the SH3–SH2–kinase domain organization is common to most cytoplasmic tyrosine kinases, namely the Src, Brk/Srm/Frk, Csk, and the Abl families (Mattsson et al., 1996; Serfas and Tyner, 2003), while the Syk/Zap family lacks an SH3 domain and has two SH2 domains and the Fes/Fer family has a single SH2 module located N-terminally to the kinase domain. Jak, Fak, and Ack cytoplasmic tyrosine kinases lack SH3 and SH2 domains. New kinases have emerged by simultaneous duplication of all the three domains together (Nars and Vihinen, 2001).

Owing to that the function of SH3, SH2, and kinase (SH1) domains has been widely investigated and reviewed (Lappalainen et al., 2008; Pawson and Scott, 2005; Pawson et al., 2001; Seet et al., 2006; Williams and Zvelebil, 2004) and since these entities are not unique for TFKs, we will not discuss them in detail, but simply make a few remarks regarding TFK-related properties.

To date there are structures for the SH3 domain of Btk (Hansson et al., 1998), Itk (Andreotti et al., 1997; Laederach et al., 2002, 2003; Severin et al., 2008), and Tec (Pursglove et al., 2002). The thermal unfolding pattern of different TFK SH3 domains has also been analyzed (Knapp et al., 1998). In contrast to most other SH3 domains, the TFK family seems to be regulated by tyrosine phosphorylation as demonstrated for Btk (Park et al., 1996) and Itk (Hao and August, 2002; Wilcox and Berg, 2003) as well as Tec and Bmx (Nore et al., 2003). In Btk, Y223 is in addition to autophosphorylation (Park et al., 1996) target also for Abl (Bäckesjö et al., 2002).

Since SH3 domains bind proline-rich stretches, it is interesting to note that Btk, Itk, and Tec have conserved PRRs in the adjacent TH domain that could bind to the SH3 domain. The Btk motif structure has been solved for BTK together with PH domain (Hyvönen and Saraste, 1997), and the entire TH domain for BMX and ITK [Protein Data Bank (PDB) codes 2ys2 and 2e61, respectively]. Bmx lacks both a typical SH3 domain and the PRR. The fact that both stretches are altered in Bmx is compatible with a functional relationship between a proline-rich, or a polyproline-like region, and the SH3 domain, as manifested in Btk (Hansson et al., 2001a,b; Laederach et al., 2002), Itk (Andreotti et al., 1997; Brazin et al., 2000; Hao and August, 2002; Márquez et al., 2003) as well as in SFKs (Moarefi et al., 1997; Sicheri et al., 1997; Wang et al., 2007; Williams et al., 1997; Xu et al., 1997). The PH and TH domains are missing from Txk.

In the case of Btk there are two PRRs in the TH domain (Smith et al., 2001; Vihinen et al., 1994a) making it possible to form multiple interactions, both intra- and intermolecular in origin (Hansson et al., 2001a,b; Laederach
et al., 2002; Okoh and Vihinen, 2002). This is not the case for Itk, which has a single proline-rich stretch, while for Txk the interaction is mainly intermolecular due to the short connecting linker (Laederach et al., 2002, 2003). However, most of these studies were performed using constructs expressing only the PRR in combination with the SH3 domain. In the single study where a full-length TFK was studied, there was no indication of TH–SH3 interactions (Márquez et al., 2003). As many multidomain proteins, including TFKs, have several overall protein folds depending, for example, on posttranslational modifications such as phosphorylation, or ligand binding, the intramolecular interactions vary. Kinases are known to have several such conformational changes due to active site loop phosphorylation, upper lobe twist because of ATP binding, and C-terminal phosphorylation in regulation of SFKs etc. Recently, Joseph et al. (2007a) presented evidence for an intramolecular cis mechanism for the phosphorylation of tyrosine 180 in the Itk SH3 domain.

NMR structural studies combined with mutational analysis demonstrated a proline-dependent conformational switch within the Itk SH2 domain. This switch regulates substrate recognition (Brazin et al., 2000). Cis–trans isomerization of a single prolyl-imide bond (D286–P287) within the SH2 domain influenced substrate recognition (Breheny et al., 2003; Mallis et al., 2002). In Btk the corresponding protein has only trans conformation (Huang et al., 2006). Originally, the BTK SH2 domain was modeled (Vihinen et al., 1994b) and used to analyze structure–function relationships along with biophysical methods for selected mutants (Mattsson et al., 2000). The structural consequences of all reported SH2 domain mutations in altogether 10 diseases have been investigated with bioinformatics methods (Lappalainen et al., 2008). Cis–trans isomerization instead seems to take place within the Btk PH domain (Yu et al., 2006) as described below.

The linker between the SH2 and kinase domain in Itk was recently shown to positively regulate catalysis (Joseph et al., 2007b) and the SH2 domain seems to be involved in substrate binding, not only in Itk, but also in Btk and Tec (Joseph et al., 2007c).

The kinase domain structure has been solved for BTK (Mao et al., 2001) and ITK (Brown et al., 2004). These domains are highly conserved in all TFK sequences.

B. Regulation of PH domain binding by phosphoinositide 3-kinase (PI3K)

The TFKs are the only tyrosine kinases having a PH domain. Likewise, the TH domain is unique to this family. While functional analyses of individual TFKs have demonstrated unique features, they also have many common characteristics. Activation of PI3K generates phosphatidylinositol-3,4,5-trisphosphate
(PIP$_3$) serving as a membrane docking site for the PH domain of TFKs. PIP$_3$ is the most negatively charged plasma membrane lipid, concentration of which can increase 40-fold within seconds after PI3K activation (Stephens et al., 1993). PIP$_3$ binding characterizes the PH domain of Btk (Manna et al., 2007; Nore et al., 2000; Rameh et al., 1997; Salim et al., 1996; Watanabe et al., 2003), Bmx (Ekman et al., 2000; Jiang et al., 2007; Qiu et al., 1998), Itk (August et al., 1997; Huang et al., 2007; Lu et al., 1998), as well as Tec (Kane and Watkins, 2005; Lachance et al., 2002; Tomlinson et al., 2004). TFKs have widely varying binding specificities and affinities for inositol compounds (Kojima et al., 1997). The differences in the binding have been investigated by modeling the structures for Bmx, Itk, and Tec PH domains (Okoh and Vihinen, 1999).

A large set of phosphoinositide-binding PH domains in various species were identified by combining bioinformatics with experimental studies (Park et al., 2008). This study confirmed the PIP$_3$-inducible binding of Btk and Tec. Altogether, 40 PI3K-regulated PH domain-containing proteins were identified in vertebrates, four in D. melanogaster, one of which was Btk29A, but none in yeast. Amino acids across the whole PH domain were found to contribute to PIP$_3$ binding. The evolutionary interpretations from the study were that PIP$_3$ regulation of PH domains has evolved several times as independent events.

PI3K can be subdivided into three classes: IA, IB; II; and III (Vanhaesebroeck et al., 1997). From an evolutionary point of view PI3K has been identified both in yeast and in plants (Herman et al., 1992; Hong and Verma, 1994; Welters et al., 1994). As PIP$_3$ can be generated via stimulation of many different receptors, including immunoreceptors, G protein-coupled receptors, as well as membrane-spanning tyrosine kinases, PI3K-induced activation of TFKs can be achieved through multiple pathways. At physiological concentrations, IP$_4$ enhances the binding of Itk’s PH domain to PIP$_3$ (Huang et al., 2007). In the case of Btk, its selective interaction with particular phosphoinositides has been addressed using biochemical and cell-biological methods (Hamman et al., 2002; Nore et al., 2000; Rameh et al., 1997; Saito et al., 2001; Salim et al., 1996; Varnai et al., 2005) as well as structure determination (Baraldi et al., 1999).

C. Regulation of TFKs through the PH domain by serine/threonine kinases

The PH domain also serves as an important region for the regulation of TFKs by serine/threonine kinases. However, the situation is quite complex with differential effects among family members. The activation of protein kinase C negatively regulates the activity of Btk (Yao et al., 1994). These and other authors demonstrated that PKC/βI constitutively interacts with Btk in vivo and that both Ca$^{2+}$-dependent and -independent forms of PKC could bind to Btk (Johannes et al., 1999; Kawakami et al., 2000; Yao et al., 1994). Btk also serves as a substrate for
PKC and its enzymatic activity is downregulated by PKC-mediated phosphorylation. However, more recent studies have reported that the key regulatory site, S180, is in fact in the TH domain (Kang et al., 2001; Venkataraman et al., 2006).

In platelets, PKCq activates Btk, while Btk negatively regulates PKCq (Crosby and Poole, 2002). Although, PKCβI and -II serve as negative regulators of Btk, the deletion of the gene encoding PKCβI and -II in mice causes a phenocopy of Btk deficiency (Leitges et al., 1996). In Itk, the situation may be different, since PKC seems to activate it (Kawakami et al., 1996).

Tec binds constitutively to PKCθ through its PH domain (Altman et al., 2004). The Bmx-induced DNA binding of Stat1 is selectively inhibited by PKCδ. The coexpression of Bmx with PKCδ-induced phosphorylation of this isoform of PKC (Saharinen et al., 1997). The interaction between PH domains and PKC seems to extend to at least some other PH domains, since the PH domain of G protein-coupled receptor kinase-2, GRK2, binds to PKC and affects the activity of this kinase (Yang et al., 2003).

Evolutionarily, PKCs are ancient and found also in plants (Bögre et al., 2003; Zegzouti et al., 2006) and in yeast (Levin et al., 1990). The same is true for phospholipase C, which appeared more than 1000 million years ago (Hirayama et al., 1995; Koyanagi et al., 1998; Tasma et al., 2008). PLCγ2 acts upstream of PKCs and serves as a substrate for Btk, which phosphorylates two tyrosines in a linker between the SH2 and SH3 domains (Humphries et al., 2004; Kim et al., 2004).

Recently, the importance of serines and PSK phosphorylation sites has been revealed from another direction. Peptidylprolyl-isomerase Pin1 targets in BTK two PH domain dipeptides, S21–P22 and S115–P116, and acts as a negative regulator. Pin1 affects the most N-terminal site in mitosis and S115–P116 during the interphase (Yu et al., 2006). Corresponding serine/threonine kinases or phosphatases have not been identified until now. Pin1 appears also in plants (Yao et al., 2001) and in yeast (Behrsin et al., 2007; Hanes et al., 1989; Lu et al., 1996).

D. Function of the TH domain

Another conserved, characteristic region of TFKs is the TH domain, and especially its N-terminal Btk motif (Smith et al., 1994b; Vihinen et al., 1994a, 1997a). While the PRR differs among TFKs, the Btk motif is highly conserved. Its only known function is to stabilize the PH domain through the conformational interaction with a Zn$^{2+}$ ion, as demonstrated for Btk (Baraldi et al., 1999; Hyvönen and Saraste, 1997). The entire TH domain is unique for TFKs. The Btk motif appears also in some other proteins. In the in-depth report of TH domain identification (Vihinen et al., 1994a) an unidentified partial protein was detected, which turned out to be Ras GTPase-activating protein (GAP) (Vihinen et al., 1997a).
As discussed above, the PRR part of the TH domain can interact with the SH3 domain in an inter- or intramolecular manner. Whether additional functions exist is not known. Missense mutations affecting patients with XLA result in a very unstable protein, which readily is degraded (Vihinen et al., 1997a). Structure for separately expressed domains in Bmx and Itk with Zn$^{2+}$ have been solved.

### E. Regulation of Btk through SH3BP5

SH3BP5 (Sab) was originally identified in mammals as a novel adaptor protein, which binds to the SH3 domain of Btk with high preference (Matsushita et al., 1998). SH3BP5 also controls negatively B-cell antigen-mediated signaling (Yamadori et al., 1999). D. melanogaster SH3BP5 (Sinka et al., 2002), also denoted Parcas, was demonstrated in a genetic screen to act as a negative regulator of Btk29A (Hamada et al., 2005).

SH3BP5 also interacts with c-Jun N-terminal kinase (JNK) (Wiltshire et al., 2002). As with c-Jun, the JNK interaction is mediated through its putative mitogen-activated protein kinase interaction motifs (KIMs) (Wiltshire et al., 2002, 2004). Active JNK and phosphorylated Sab are colocalized and associated with mitochondria (Wiltshire et al., 2002). These findings suggest a role in crosstalk between Btk and JNK signaling pathways (Wiltshire et al., 2002, 2004).

A SH3BP5-like (SH3BP5L) protein shares sequence similarity with SH3BP5 (Strausberg et al., 2002). No biological function has been characterized for it yet.

### F. Mutations in X-linked agammaglobulinemia as a tool to study the function of residues

XLA, being the protein kinase with the largest number of disease-related mutation among all human protein kinases (Ortutay et al., 2005), has allowed us and others to investigate the significance and structure–function relationships of several regions and amino acids in TFKs and other kinases. Missense and other in-frame mutations have been instrumental in these studies. Due to the conservation of TFKs along the entire sequence it is relatively easy to interpolate information for one family to others as in the case of JAK3 (Notarangelo et al., 2001; Vihinen et al., 2000). The importance of such data is not limited to TFKs, since when the naturally occurring, disease-causing mutation R28C in the Btk PH domain, in both in man and mice, was introduced to the Akt kinase, a stable but inactive protein was obtained (Lehnes et al., 2007; Sable et al., 1998; Stoica et al., 2003).

Since 1995 mutations in the BTK gene have been collected. The BTKbase (http://bioinf.uta.fi/BTKbase) database is freely available and has served as a model for some 130 additional mutation databases, mainly for
immunodeficiencies (Piirilä et al., 2006). BTKbase has constantly grown from 188 cases to the current number of 1096 patients (Lindvall et al., 2005; Väliaho et al., 2006; Vihinen et al., 1995a,b, 1996, 1997b, 1998, 1999, 2001). Experimental and modeled structures for BTK domains have extensively been used to explain protein structure–function relationships and consequences of mutations (Holinski-Feder et al., 1998; Jin et al., 1995; Korpi et al., 2000; Lindvall et al., 2005; Maniar et al., 1995; Mao et al., 2001; Mattsson et al., 2000; Okoh et al., 2002; Speletas et al., 2001; Väliaho et al., 2006; Vihinen et al., 1994b,c, 1995b; Vorechovsky et al., 1995, 1997; Zhu et al., 1994). These studies have allowed us to provide putative functional and/or structural explanation for all XLA-causing mutations. The quality of the structural predictions was retrospectively assessed and found to be very good (Khan and Vihinen, submitted for publication).

Mutations have also been detected in the BTK promoter region affecting a highly conserved binding site for Ets family transcription factors (Holinski-Feder et al., 1998). Btk has been reported to activate NF-κB signaling (Bajpai et al., 2000; Petro et al., 2000). Recently Btk was shown to autoregulate its promoter in a positive fashion (Yu et al., 2008). Interestingly, also Bmx, Itk, and Tec seem to be positively regulated by NF-κB (Yu et al., 2008, manuscript in preparation).

IV. THE ANCESTRY OF TFKs

A. Collecting the sequences

Multiple blastp (Altschul et al., 1997) searches were made in an iterative manner against GenBank nonredundant protein database to identify TFK sequences. In the first step, human TFK protein sequences were used as query and the results were examined by distant tree representation with the online tool provided by the National Center for Biotechnology Information (NCBI). All the detected sequences were carefully checked and only full-length entries were accepted. When sequence variants appeared, they were aligned to the query sequence and the one closest to the query with the longest sequence was selected. From each search only those sequences, which were closest to the human query were included to the result set, that is, only the orthologs were included. Sequences for which the orthology was not so clear were collected separately. Additional search was made by using the fruit fly Btk29A sequence as the query.

TFK members were identified also from some incomplete genome data-sets. Unannotated TFK sequences were searched from Xenopus laevis (Bowes et al., 2008), Xenopus tropicalis (Bowes et al., 2008), Takifugu rubripes (Aparicio et al., 2002), and Danio rerio (zebrafish) (Sprague et al., 2008) genomes. Several tblastn (Altschul et al., 1997) searches were performed using TFK protein
sequences from *Gallus gallus* (chicken) as queries. An obviously missing Btk sequence from *Macaca mulatta* was identified by using a tblastn search against Macaca mRNA database (Gibbs et al., 2007). A likely mistranslated mRNA was detected with a frameshift caused by a possible sequencing error. The sequence was reconstructed by translating the mRNA in six reading frames and by aligning the protein sequences with the human BTK protein sequence. Partial sequences, such as those for Tec and Txk in *Sus scrofa* (wild boar), were excluded.

**B. Aligning the sequences**

The identified TFK sequences were divided into seven groups: the chordate-specific Bmx, Brk, Itk, Tec, and Txk groups and the insect-specific Btk29A-related sequences. The seventh group is for three sequences equally distant from all the others. This outgroup contained sequences BAD52302 from *Eptatretus burgeri* (hagfish), XP_001745298 from *Monosiga brevicollis* (choanoflagellate), and AAP82507 from *Suberites domuncula* (sponge). The six protein groups were aligned individually and finally combined with the outgroup sequences. In this way the internal conservation of the protein groups was better preserved in the final alignment.

**C. Phylogenetic analysis**

Based on the multiple sequence alignment, a bootstrap analysis was performed using maximum parsimony as criteria for searching the optimal tree (Fig. 3.1). The six protein groups are clearly separated on the tree, and we can draw the phylogeny of these proteins as follows. The ancestor of all the TFKs was present in early eukaryotes prior to the formation of metazoans. The sequences from *S. domuncula* and *M. brevicollis* are orthologs of the ancestor. After the divergence of deuterostomia and protostomia the descendants of the ancestor further diverged. In protostomia, now insects, TFKs developed in the form of the Btk29A protein group.

In deuterostomia a descendant of the single gene became the ancestor for the five chordata-specific protein groups. The TFK in *E. brugeri* is a direct ascendant to that. After the formation of craniata, but before the formation of vertebrata, the ancestor went through multiple duplications. First, it was divided into the Btk/Bmx and Tec/Txk/Itk groups. Then both groups duplicated until all the five protein groups appeared. These events took place before the emergence of vertebrates. The lack of sequences in fishes and some other genomes is likely due to deletion events rather than duplications after the emergence of the vertebrates, since all the sequences within the groups are more similar to each other than to any of the fish or frog sequences.
Figure 3.1. Phylogenetic relationship of the TFK sequences. The six major protein groups can be clearly distinguished. Bootstrap analysis was carried out using maximum parsimony criteria and 100 replicates with PAUP* (Swofford, 2003). Bootstrap values are shown at the nodes. Sequence labels contain name of the species and NCBI Entrez accession number for the protein sequence. Btk sequence for *Macaca mulatta* was reconstructed by transcribing the mRNA record XR_011285.1 and correcting the disrupted frameshift.
Another view on evolution is based upon the analysis of genomic organization. Recently the amphioxus, *Branchiostoma floridae*, genome was published (Putnam *et al.*, 2008). The authors partly reconstructed the genomic organization of the last common chordate ancestor and described two genome-wide duplications and subsequent reorganizations in the vertebrate lineage. Interestingly, number 8 of the 17 reconstructed ancestral chordate linkage groups contains regions corresponding to the location of all TFKs in the human genome.

Since we identified only a single frog TFK, the genomic region was analyzed. According to Xenbase (http://www.xenbase.org/), *Cyfip2* and *Med7* genes are both located in the vicinity of the *Tec* gene. This is surprising since both these genes are in close proximity to the *Itk* gene in zebrafish, mouse and man. The *Itk* and *Tec* genes in these three species are on different chromosomes. The *Txk* gene, which is absent from the zebrafish, is in very close proximity to the *Tec* gene in humans and mice. However, there is no doubt from the sequence alignment that the frog TFK should be classified as Tec. It is possible that a recombination event in an ancestor have transferred the *Tec* gene from its original location into the position of an *Itk* homolog, which was simultaneously lost. An alternative explanation is that the gene has evolved so that it is currently more closer to *Tec* than to *Itk*. *X. laevis* Tec has similarity to Tec from other species throughout the sequence, suggesting that if a recombination occurred, the entire gene was replaced.

**D. The origin of the Btk-specific PH domain loop in amniotes**

Many Btk sequences contain a characteristic loop of 20 residues between amino acids 78 and 98 (numbering according to BTK). It is present only in the mammalian and bird Btk orthologs, not in the other TFKs. Exon 4 in these genes encodes residues 81–103 (human BTK numbering). We assume that the loop emerged by the insertion of exon 4 into the *Btk* gene. This insertion was tolerated, because the loop is on the surface of the protein (Fig. 3.2). The termini are close to each other in the three-dimensional space as in many protein domains, which appear in several proteins surrounded by different structural regions. In the loop region, there are in BTKbase, five frameshift mutations, which lead to premature stop codon and truncated product, three nonsense mutations, and a duplication, which causes a frameshift and premature termination. These mutations are disease causing, because they produce a nonfunctional, truncated protein. There is just one missense mutation, F98V substitution. The presence of a single, disease-causing missense mutation is not very informative, since this stretch is rather short and does not contain any hotspots for mutations, such as repetitive sequences or CpG dinucleotides. The loop is highly conserved.
There are just minor variations in the sequence alignment, which suggests that selection pressure has maintained the sequence. Whether the underlying mechanism is functional or has a different origin is not clear.

There is also a missense SNP in the region, rs56035945 with R82K alteration, without known phenotype. Chicken, platypus and opossum Btks have a lysine at that site while there is arginine in the higher mammals. Arginine and lysine are both basic residues and frequently substituted by each other in protein families.

E. N-terminal regions of insect TFKs

*D. melanogaster* has three alternatively spliced mRNA forms, whereas a single form appears in other insects. The three mRNA variants code for two alternative protein products, because two of them differ just in their 5′-untranslated regions. The difference in the variants is that the longer form encompasses exons 1–5 and 9–16, while the shorter form contains exons 5–16. The longer protein variant aligns with that from *D. pseudoobscura*, while the shorter one with the proteins from *Anopheles gambiae* and *Culex pipiens quinquefasciatus*. Also other dipteran genomes might have alternatively spliced forms.
The N-terminus in insect proteins does not align at all with other TFK proteins before the beginning of the SH3 domain from whereon they align very well. Since this region is not conserved at all, and the whole N-terminus is truncated in the *Aedes aegypti* sequence, its specific function remains elusive.

V. SH3BP5—A CONSERVED NEGATIVE REGULATOR OF TFKs

The SH3-binding protein 5 interacts with Btk and MAP kinases. The ortholog sequences of human SH3BP5 protein were collected in a similar way as for TFKs. As there were closely related paralogs called SH3BP5L in several genomes they were included to the analysis.

In the tree, there are several protein groups (Fig. 3.3). An ortholog appears also in *M. brevicollis*, but does not fit to any of the groups. The insect and vertebrate groups for both SH3BP5 and SH3BP5L can be clearly distinguished along with a group in nematodes. The relationship of these big groups is more difficult to resolve. The simplest explanation for the tree is that there was a single SH3BP5 gene in the ancestors of eukaryotes and a duplication appeared very early. Both the duplicons were then preserved in insect as well as vertebrate lineages. The nematode group does not fit easily to this explanation. The role of both SH3BP5 and SH3BP5L proteins has to be significant, since they can be identified from all the important branches of eukaryotes. The detailed function of these proteins remains to be elucidated.

VI. THE ORIGIN OF PHOSPHOTYROSINE SIGNALING AND THE ROLE OF CYTOPLASMIC TYROSINE KINASES

The evolution of phosphotyrosine signaling suggests that more than 600 million years ago there was a common ancestor for the unicellular choanoflagellates and for multicellular metazoans, which had already developed this ability (King and Carroll, 2001; King *et al.*, 2008; Peterson and Butterfield, 2005; Pincus *et al.*, 2008). In some species, such as in yeast, tyrosine phosphorylation appears at a very low level, most likely due to promiscuity of serine/threonine kinases (Schieve *et al.*, 1986). In a recent report, Pincus *et al.* (2008) suggest that phosphatases and SH2 domains appeared first, whereas the enzymatic activity of tyrosine kinases developed later. The emergence of specific proteins resulted in the expansion of proteins and domains in cellular signaling. One third of all domains found in combination with SH2 domains in choanoflagellates are unique while 38% are shared with metazoans.
Proteins active in tyrosine kinase-related signaling are quite abundant in choanoflagellates (King et al., 2008). While most proteins in serine–threonine signaling pathways are common between metazoans and choanoflagellates, the opposite is true for tyrosine phosphorylation and several other intracellular signaling pathways, including many transcription factors.

Figure 3.3. Phylogenetic relationship of the SH3-binding protein 5 and related sequences. Groups of SH3BP5 and SH3BP5L proteins from vertebrates and insects can be distinguished in addition to a distinct group of SH3BP5 proteins from nematodes. Further relationship of these groups is not clear. The analysis was carried out as described in Fig. 3.1.

Proteins active in tyrosine kinase-related signaling are quite abundant in choanoflagellates (King et al., 2008). While most proteins in serine–threonine signaling pathways are common between metazoans and choanoflagellates, the opposite is true for tyrosine phosphorylation and several other intracellular signaling pathways, including many transcription factors.
Among choanoflagellates, SFKs and C-terminal Src kinase (Csk) were first reported in M. ovata (Segawa et al., 2006). The M. ovata Src has transforming capability but is not negatively regulated by Csk. Biochemical characterization of the M. brevicollis Src and Csk indicated that the putative, regulatory C-terminal tyrosine is not phosphorylated (Li et al., 2008).

Our report is the first to demonstrate the existence of a TFK in M. brevicollis, which has estimated to have 128 tyrosine kinase genes (King et al., 2008). Although the role of the M. brevicollis TFK is unknown, its mere existence clearly suggests that it is functionally active in a unicellular organism. Since all domains of TFKs are conserved in this protein, it is possible that this kinase is already regulated by SFKs, PI3K, and PKC, like the metazoan counterparts. However, given the lack of Csk-induced control of SFKs in M. brevicollis as well as in M. ovata, it is equally possible that the regulation differs. Choanoflagellates also encode an SH3BP5-related molecule, which has not been functionally characterized. Thus, it is too early say whether this molecule suppresses the corresponding TFK. Functional studies will be needed to resolve this issue as well as the possibility that the choanoflagellate TFK can substitute for the loss of TFKs in metazoan cells.

Our study of TFKs reveals that these enzymes are ancient and their ancestor appeared already in choanoflagellates. TFK members are regulated by several proteins and they control numerous signaling pathways. More studies will be needed to investigate how the pathways in which TFKs currently participate originally obtained this property.

**Acknowledgments**

This work was supported by The Swedish Cancer Fund, The Wallenberg Foundation, the Swedish Science Council, the Stockholm County Council (research grant ALF-projektmedel), the European Union grant FP7-HEALTH-F2-2008-201549, the Medical Research Fund of Tampere University Hospital, and Academy of Finland.

**References**


3. Phylogeny of Tec Family Kinases


3. Phylogeny of Tec Family Kinases


