A mysterious family of calcium binding proteins from parasitic worms
Charlotte M. Thomas\textsuperscript{1,2} and David J. Timson\textsuperscript{1,3}\textsuperscript{*}

\textsuperscript{1} School of Biological Sciences, Queen's University Belfast, Medical Biology Centre, 97 Lisburn Road, Belfast, BT9 7BL, UK.
\textsuperscript{2} Institute for Global Food Security, Queen's University Belfast, 18-30 Malone Road, Belfast, BT9 5BN, UK.
\textsuperscript{3} School of Pharmacy and Biomolecular Sciences, University of Brighton, Huxley Building, Lewes Road, Brighton, BN2 4GJ, UK.

\textsuperscript{*}Author to whom correspondence should be addressed at School of Pharmacy and Biomolecular Sciences, University of Brighton, Huxley Building, Lewes Road, Brighton, BN2 4GJ, UK.
Telephone \textsuperscript{+44(0)1273 641623}
Fax \textsuperscript{+44(0)1273 642090}
Email \textsuperscript{d.timson@brighton.ac.uk}
Abstract
There is a family of proteins from parasitic worms which combine N-terminal EF-hand domains with C-terminal dynein light chain-like domains. Data are accumulating on the biochemistry and cell biology of these proteins. However, little is known about their functions in vivo. *Schistosoma mansoni* expresses 13 family members (SmTAL1-SmTAL13). Three of these (SmTAL1, SmTAL2 and SmTAL3) have been subjected to biochemical analysis which demonstrated that they have different molecular properties. Although their overall folds are predicted to be similar, small changes in the EF-hand domains result in differences in their ion binding properties. While SmTAL1 and SmTAL2 are able to bind calcium (and some other) ions, SmTAL3 appears to be unable to bind any divalent cations. Similar biochemical diversity has been seen in the CaBP proteins from *Fasciola hepatica*. Four family members are known (FhCaBP1-4). All of these bind to calcium ions. However, FhCaBP4 dimerises in the presence of calcium ions, FhCaBP3 dimerises in the absence of calcium ions and FhCaBP2 dimerises regardless of the prevailing calcium ion concentration. In both the SmTAL and FhCaBP families, the proteins also differ in their ability to bind calmodulin antagonists and related drugs. Interestingly, SmTAL1 interacts with praziquantel (the drug of choice for treating schistosomiasis). The pharmacological significance (if any) of this finding is unknown.
Introduction: calcium signalling in cells

Changes in the cellular calcium ion concentration acts as a signal in a wide range of organisms from throughout the tree of life [1]. Calcium ions are well-suited to this task since, of the commonly occurring inorganic ions in living systems, solvating water molecules have the highest association and dissociation rates [2, 3]. This means that these water molecules are rapidly exchanged between the ion and the water. Since interaction proteins generally requires the removal of some, or all, of these water molecules binding to proteins can also be rapid. By the same argument, removal of calcium ions from proteins and their return to aqueous solution can also be rapid. Thus, calcium binding and unbinding enables the rapid detection and termination of signals.

Signalling by calcium ions necessarily requires that the ions are recognised by proteins which then respond, normally through conformational changes which alter interactions with other proteins [4, 5]. Many cells express a variety of these calcium sensors, and the best characterised is calmodulin. This protein has two globular domains linked by an extended alpha-helical segment [6, 7]. Each of the two globular domains contains two EF-hand motifs which are capable of binding calcium and other divalent ions [5, 8]. The EF-hand motif folds into a tight turn which, through a combination of side chain and backbone functional groups, is able to coordinate a calcium ion. Considerable work has been done to define the consensus residues required for calcium ion binding [4]. These residues include ones with carboxylate or hydroxyl side chains involved in the coordination of the ion and also glycines to enable the tight turns required to fold around the ion [4]. In response to calcium binding, calmodulin undergoes a conformational change in which the protein's surface becomes more hydrophobic [9, 10]. This change in surface properties alters how the protein interacts with other molecules. In many cases, this change enables the binding to partner proteins, but there are also cases where it reduces the affinity. Calmodulin is implicated in regulating a wide variety of cellular processes including muscle contraction, calcium homeostasis and cytokinesis [11-13].

Calcium sensors such as calmodulin have attracted interest as drug targets. A number of neuro-active compounds have been shown to act, at least in part, through the antagonism of calmodulin’s signalling to other proteins. These drugs include trifluoperazine (TFP) and chlorpromazine (CPZ); these are both used in the treatment of schizophrenia and other psychiatric diseases [14-17]. There has also been some interest in exploiting the antagonism of calmodulin in the development of novel anti-cancer treatments [18]. These compounds act by binding to hydrophobic patches and crevices on the surface of the calcium-calmodulin complex, blocking the protein's interactions with other biomolecules [10, 19].

An unusual family of calcium-binding proteins in parasitic worms

Parasitic worms cause a significant disease burden on humanity [20]. This is most keenly felt in the developing world. There is also a substantial effect on global agriculture and food security with losses of livestock production due to worm infections estimated at billions of dollars per annum. While there are a number of safe and effective drugs for the treatment of parasitic worm infections, resistance to these is emerging and likely to become more prevalent as selective pressures drive evolution [21]. There is limited research into alternative treatments for parasitic worm infections and there is also a lack of knowledge about the fundamental biochemistry of these organisms. Consequently, there is a need to understand systems such as calcium signalling in these organisms in order to build up the knowledge base that may, one day, be used in novel drug discovery.
Many species of parasitic worms, especially those in the class Trematoda (flukes) express members of a family of unusual calcium-binding proteins. These proteins combine an N-terminal domain with two EF-hand motifs and a C-terminal domain which has high similarity to dynein light chains (DLC-like domain). Dynein light chains are accessory subunits of the microtubule motor dynein [22, 23]. They are typically small proteins (~8-10 kDa) which have the capacity to homodimerise [24]. Structurally, they consist of a largely β-sheet structure [24-26]. Dynein light chains interact with a wide range of diverse proteins which may include some “cargo” molecules to be carried by the dynein motor [26, 27]. The two domains are joined by a flexible linker sequence (Figure 1). No similar proteins have been identified in any other group of organisms. This combination of their uniqueness to parasitic worms and their likely involvement in calcium signalling make them attractive targets for further investigation as potential drug targets.

Interestingly, trematodes appear to express a relatively large number of these proteins (Table 1). The common liver fluke (Fasciola hepatica) is known to express at least four isoforms [28-31], and preliminary analysis of the draft genome suggests that the total is more than ten [32] (CMT, unpublished data). The blood fluke Schistosoma mansoni expresses 13 isoforms [33] and it seems likely that other species from the same genus express similar numbers [34]. Examples from other trematode species have also been identified (Table 1). However, there have not been systematic surveys of isoform diversity in these species to date.

Biochemical analysis of these proteins has revealed differences in their properties. Four isoforms from F. hepatica have been characterised to date. FH22 (FhCaBP1), FhCaBP2, FhCaBP3 and FhCaBP4 all bind to calcium ions [28-31]. In the case of FhCaBP4, this interaction promotes dimerisation of the protein [30]. In contrast, dimerisation is favoured in the calcium-free state of FhCaBP3 [29] and FhCaBP2 dimerisation appears to be unaffected by the presence or absence of calcium ions [28]. All four proteins are predicted to have similar structures by molecular modelling [28-30] (DJT unpublished data) (Figure 1). To date, no experimental structures have been released. The ion binding properties of FhCaBP1 have not been investigated in detail. FhCaBP2, FhCaBP3 and FhCaBP4 show different properties. While all three bind calcium and manganese ions, they differ in the other divalent ions which can be recognised (Table 1) [28-30]. They also can be distinguished pharmacologically with FhCaBP2 and FhCaBP3 binding to a range of calmodulin antagonists including trifluoperazine, chlorpromazine and N-(6-aminohexyl)-5-chloro-1-naphthalen sulphonamide (W7) [28, 29]. In contrast FhCaBP4 binds to W7 but not trifluoperazine [30].

A more detailed study of FhCaBP2 confirmed that calcium ion binding occurs through the N-terminal domain (which contains the two EF-hands) [28]. Of these two EF-hands the second is more important in binding to the ion: mutation of a key aspartate residue (Asp-58) in this motif abolished calcium ion binding. Both the isolated N- and C-terminal domains are capable of dimerising, strongly suggesting that dimers of the intact proteins dimerise along their long axis involving contacts from both domains [28]. In some isoforms it has been demonstrated that calcium-ion binding increases the surface hydrophobicity of the molecule (analogously to calmodulin) [28, 30]. However, it is currently unclear if (and how) calcium ion binding by the EF-hand can affect the DLC-like domain.
Biochemical analysis of members of the protein family from *S. mansoni* (known in this species as tegumental allergen-like proteins, TALs due to their ability to induce IgE-mediated immune responses [35]) have also revealed differences. Once again, the predicted structures are similar [36] (DJT, unpublished data). The crystallisation of one isoform (Sm21.7 or SmTAL2) has been reported, but the structure is not yet available. Interestingly this protein spontaneously degraded under the conditions of crystallisation such that only the DLC-like domain remained [37]. Ion binding properties of these TAL proteins varies widely. SmTAL1 interacts with calcium, manganese strontium and nickel (II) ions, whereas SmTAL2 only interacts with calcium and manganese only [36]. However, there is no evidence that SmTAL3 binds to calcium or any other divalent cation [36, 38]. These experimental findings are consistent with the predicted structure of the second EF-hand in SmTAL3 which deviates from the consensus sequence in several places [4, 36]. All three of these proteins are capable of dimerising and calcium ions have no effect of the detected extent of dimerisation [36]. To date, no detailed biochemical studies on the remaining family members have been reported. SmTAL1, SmTAL2 and SmTAL3 can be distinguished pharmacologically: each has distinct drug binding properties [36]. Interestingly, SmTAL1 interacts with praziquantel, the current drug of choice for the treatment and prevention of infection with schistosomes [36, 39]. It is not clear if this interaction is pharmacologically important. However, it is well-established that praziquantel acts through the dysregulation of calcium homeostasis, resulting in rapid influx of calcium ions. This uncontrolled rise in the intracellular calcium ion concentrations causes the disruption of a variety of processes most importantly resulting in the muscle contraction and paralysis of the worm [40]. The precise, molecular mechanism of action of praziquantel remains unknown. It is tantalising to suggest that its interaction with SmTAL1 may be important in this mechanism. However, there is currently no evidence to support this postulate.

**Mysteries associated with this protein family**

There remain many mysteries about this protein family. Not all biochemical questions have been resolved. While many differences in ion binding and oligomerisation properties have been discovered, the underlying structural causes of these are not yet defined. The dimerization interfaces have not been mapped and further studies on the flexibility of the linker which joins the two domains would also be desirable.

The subtle differences in biochemical properties suggest that they perform discrete cellular roles. However, it is not clear why so many isoforms are required by parasitic worms. It is possible that there is some overlap or redundancy in function or that some proteins are only used in specific life-cycle stages. (It should be noted that the conditions of the life cycle stages vary considerably. For example, mature flukes exist largely inside mammals at the controlled temperatures prevailing there. In contrast, free living stages are subject to greater temperature fluctuations and typically lower temperatures than those found inside mammalian bodies.) The role of manganese binding is also unclear. To date, no isoform has been identified which binds calcium ions, but not manganese ions. This may be an experimental artefact or may have potential physiological significance. There are few examples of manganese ion based signalling but the possibility remains that it may be occurring in these systems.

The biggest mystery surrounding this protein family is their roles and functions. The different biochemical properties suggest that they play a variety of roles in the parasite [41]. However, there are currently no clear indications what these roles may be. An early study on SmTAL3 (Sm20.8)
demonstrated that the protein can be isolated in a large, multi-protein complex which contained dynein [42]. The presence of DLC-like domains may mean that members of this proteins family can interact directly with dynein. However, it should be noted that breakage of cells and extraction of proteins would necessarily result in the mixing of proteins and so there remains the possibility that SmTAL3's presence in this complex resulted from a non-physiological association of the DLC-like domain with dynein. Localisation studies on family members from a variety of species have demonstrated that the proteins are often found in the tegument [43-45]. This syncytium is located on the surface of the worm and is intimately involved in host-parasite recognition. It is a highly dynamic and metabolically active structure and it seems likely that it has a strong requirement for a variety of tightly regulated processes, including those mediated by dynein, for example vesicular transport [46]. Therefore, there may be a role for family members in either directly regulating these cytoskeletal processes or in the maintenance of calcium homeostasis in the tegument. Mapping the signalling pathways that this group of proteins participate in would also be likely to provide important functional data.

There is a clear need to learn more about the roles of these proteins. Greater understanding of these would result from the identification of protein (or other biomacromolecule) binding partners. The current lack of good cell culture models for many parasitic worms combined with difficulties in maintaining many life cycle stages in the laboratory hinder in vivo studies. In this context it may be useful to establish simple “model systems” in which key components are expressed in simple, more genetically tractable organisms such as the budding yeast Saccharomyces cerevisiae. It may also be possible to study the effects of “knocking down” the expression of selected family members using RNAi. This technique is now well-established in many parasitic worms (see, for example [47-51]). In the case of these protein families, issues may arise due to the highly similar nature of their sequences and the possibility that some of the proteins may have long cellular half-lives (and, thus, knock-down of mRNA levels may take some time to feed through to protein levels). Nevertheless, the recent successful use of mRNA to knock down (at the mRNA and protein levels), selectively calmodulin isoforms in F. hepatica suggests that this approach may be possible [52]. If so, careful identification of likely phenotypic read-outs will be important to maximise the utility of the experiments.

This protein family are interesting at the biochemical level. Despite their overall sequence and predicted structural similarity, they demonstrate biochemical differences which are likely to be reflected in functional differences. They can be differentiated pharmacologically suggesting that it would be possible to design compounds which recognise one isofrom or a closely related group of isoforms. Thus, if it is subsequently shown that this mysterious group of proteins are involved in vital processes in parasitic worms, they may become a very attractive target for the design of drug-like molecules which target the pathogen but have no likely binding partner in the host. In order to get to this point, many more biochemical, genetic and physiological studies will be required.

**Acknowledgements**

CMT thanks the Department of Employment and Learning, Northern Ireland (DELNI, UK) for a PhD studentship. DJT wishes to thank Drs Liz Hoey and Alan Trudgett (both recently retired from Queen’s University, Belfast) for introducing him to helminth biology and biochemistry.
Figure legend

Figure 1: **Predicted structure of a typical member of this protein family.** A schematic diagram of the three regions of the protein (top) with the residues in each region shown for representative family members. (Note that FhCaBP3 has an extended, unstructured region at the N-terminus.) Domain boundaries were estimated by eye using previously published molecular models. The modelled structure of FhCaBP4 [30] shown (bottom) in the same colour scheme. This protein is shown bound to one calcium ion in the second EF-hand.
Table 1: A summary of the known properties of family members

<table>
<thead>
<tr>
<th>Species</th>
<th>Protein</th>
<th>Alternative name</th>
<th>Ion binding</th>
<th>Drug binding</th>
<th>Dimerization</th>
<th>Localisation</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fasciola hepatica</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FhCaBP1</td>
<td>Fh22</td>
<td></td>
<td>Ca(^{2+})</td>
<td>??</td>
<td>??</td>
<td>Excretory-secretory extract</td>
<td>[31]</td>
</tr>
<tr>
<td>FhCaBP2</td>
<td></td>
<td></td>
<td>Ca(^{2+}), Mn(^{2+})</td>
<td>TFP, CPZ, W7, ThA</td>
<td>Yes +/-Ca(^{2+})</td>
<td>??</td>
<td>[28]</td>
</tr>
<tr>
<td>FhCaBP3</td>
<td></td>
<td></td>
<td>Ca(^{2+}), Mn(^{2+})</td>
<td>TFP, CPZ, PZQ</td>
<td>Yes -Ca(^{2+})</td>
<td>??</td>
<td>[29]</td>
</tr>
<tr>
<td>FhCaBP4</td>
<td></td>
<td></td>
<td>Ca(^{2+}), Mn(^{2+}), Ba(^{2+}), Sn(^{2+})</td>
<td>W7</td>
<td>Yes +Ca(^{2+})</td>
<td>??</td>
<td>[30]</td>
</tr>
<tr>
<td><strong>Fasciola gigantica</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FgCaBP1</td>
<td></td>
<td></td>
<td>Ca(^{2+})</td>
<td>??</td>
<td>??</td>
<td>Tegument</td>
<td>[43, 44]</td>
</tr>
<tr>
<td>FgCaBP2</td>
<td></td>
<td></td>
<td>Ca(^{2+})</td>
<td>??</td>
<td>??</td>
<td>??</td>
<td>[43]</td>
</tr>
<tr>
<td>FgCaBP3</td>
<td></td>
<td></td>
<td>Ca(^{2+})</td>
<td>??</td>
<td>??</td>
<td>Tegument</td>
<td>[43]</td>
</tr>
<tr>
<td>FgCaBP4</td>
<td></td>
<td></td>
<td>Ca(^{2+})</td>
<td>??</td>
<td>??</td>
<td>Tegument; lining of the seminal vesicle</td>
<td>[43]</td>
</tr>
<tr>
<td><strong>Schistosoma mansoni</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SmTAL1</td>
<td>Sm22.6</td>
<td></td>
<td>Ca(^{2+}), Mn(^{2+}), Ni(^{2+}), Sn(^{2+})</td>
<td>PZQ, CPZ, W7, TFP</td>
<td>Yes +/-Ca(^{2+})</td>
<td>Tegument</td>
<td>[36, 53]</td>
</tr>
<tr>
<td>SmTAL2</td>
<td>Sm21.7</td>
<td></td>
<td>Ca(^{2+}), Mn(^{2+}), Cd(^{2+}), Mg(^{2+})</td>
<td>W7</td>
<td>Yes +/-Ca(^{2+})</td>
<td>??</td>
<td>[36, 54]</td>
</tr>
<tr>
<td>SmTAL3</td>
<td>Sm20.8</td>
<td>None</td>
<td></td>
<td>TFP, ThA</td>
<td>Yes +/-Ca(^{2+})</td>
<td>Tegument</td>
<td>[36, 38]</td>
</tr>
<tr>
<td>SmTAL4</td>
<td></td>
<td></td>
<td></td>
<td>??</td>
<td>??</td>
<td>Tegument of the cercarial tail</td>
<td>[33]</td>
</tr>
<tr>
<td><strong>Schistosoma japonicum</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SjTP22.4</td>
<td></td>
<td></td>
<td>Ca(^{2+})</td>
<td>??</td>
<td>??</td>
<td>Tegument</td>
<td>[55]</td>
</tr>
<tr>
<td>Sj22.6</td>
<td></td>
<td></td>
<td></td>
<td>??</td>
<td>??</td>
<td>Apical cytoplasm</td>
<td>[56-58]</td>
</tr>
<tr>
<td>Species</td>
<td>Protein</td>
<td>Ion</td>
<td>Other</td>
<td>Tissue</td>
<td>Reference</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------</td>
<td>-----------------------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Clonorchis sinensis</em></td>
<td>CsTegu21.6</td>
<td>??</td>
<td>??</td>
<td>Tegument</td>
<td>[59]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CsTP31.8&lt;sup&gt;e&lt;/sup&gt;</td>
<td>??</td>
<td>??</td>
<td>Tegument</td>
<td>[60]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Opisthorchis viverrini</em></td>
<td>OvCaBP</td>
<td>Ca&lt;sup&gt;2+&lt;/sup&gt;</td>
<td>??</td>
<td>Gut epithelium, miracidia in eggs</td>
<td>[45]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Not all ions and drugs have been tested with all proteins. Listed here are those for which experimental evidence in favour of binding has been obtained. Therefore, absence from the list does not, necessarily, mean that the protein does not bind a particular ion or drug.

<sup>b</sup> TPZ, trifluoperazine; CPZ, chlorpromazine; ThA, thiamylal; PZQ, praziquantel

<sup>c</sup> ?? indicates something which is not currently known

<sup>d</sup> Inferred from sequence alignment

<sup>e</sup> The protein is approximately 10 kDa larger than other family members and must, therefore, have additional domains/regions compared to the structure illustrated in Figure 1.
References
and folding of LC8, a highly conserved light chain of cytoplasmic dynein. Biochemistry. 40, 1596-1605


X. Wang, Chen, W., Tian, Y., Huang, Y., Li, X. and Yu, X. (2014) RNAi-mediated silencing of enolase confirms its biological importance in Clonorchis sinensis. Parasitol Res. 113, 1451-1458


