Sarcocystis neurona Protein Kinases: Computational Identification, Evolutionary Analysis and Putative Inhibitor Docking

Edwin K. Murungi 1 & Henry M. Kariithi 2

1 Department of Biochemistry and Molecular Biology, Egerton University, P.O. Box 536, 20115, Njoro, Kenya; eddkimm@gmail.com
2 Biotechnology Research Institute, Kenya Agricultural and Livestock Research Organization, P.O Box 57811, 00200, Kaptagat Rd, Loresho, Nairobi, Kenya; henry.kariithi@kalro.org

* Correspondence: eddkimm@gmail.com; Tel: (+254)-789-716059

Abstract: The apicomplexan parasite, Sarcocystis neurona causes the degenerative neurological equine protozoal myeloencephalitis (EPM) disease of horses. Due to its host range expansion, S. neurona is an emerging threat that requires close monitoring. In apicomplexans, protein kinases (PKs) have been implicated in a myriad of critical functions such as host cell invasion, cell cycle progression and host immune responses evasion. Here, we used various bioinformatics methods to define the kinome of S. neurona and phylogenetic relatedness of its PKs to other apicomplexans. Further, three-dimensional (3D) homology models for selected S. neurona putative PKs were constructed and evaluated for inhibitor docking. We identified 92 putative PKs clustering within the AGC, CAMK, CK1, CMGC, STE, TKL, aPK and OPK groups. Although containing the universally conserved PKA (AGC group), S. neurona kinome was devoid of PKB and PKC, but contained the six apicomplexan conserved CDPKs (CAMK group). The OPK group was represented by ROPKs 19A, 27, 30, 33, 35 and 37, but was devoid of the virulence-associated ROPKs 5, 6, 18 and 38. Two out of the three S. neurona CK1 enzymes had high sequence similarities to T. gondii TgCK1-α and TgCK1-β and the Plasmodium PfCK1. Docking of four inhibitors onto homology models of putative ROP27 and PKA indicated that inhibition of S. neurona PKs is feasible, but needs to be experimentally tested. The essentiality of apicomplexan PKs makes the elucidation of S. neurona kinome a key milestone for development of novel therapeutics for EPM.

Keywords: Sarcocystis neurona; EPM; Docking; Apicomplexans; Phylogeny; Homology modeling.

1. Introduction

Equine protozoal myeloencephalitis (EPM) is an infectious, progressive, degenerative neurological disease of horses caused by the apicomplexan parasite, Sarcocystis neurona [1]. To complete its life cycle, this heteroxenous parasite requires a reservoir host (i.e. opossums; Didelphis virginiana, D. albiventris) and an aberrant (horses) or intermediate host (cats, skunks, raccoons and sea otters) [2]. Opossums become infected upon ingestion of sarcocysts containing hundreds of bradyzoites, which undergo gametogony that sporulate into mature oocysts, which are then shed in the faeces. Upon ingestion by the intermediate or aberrant hosts, the oocysts release the environmentally resistant sporozoites, which chronically parasitize the neural and inflammatory cells of the host’s central nervous system (CNS). Clinical EPM symptoms depend on which part of the CNS is parasitized; parasitization generally results in abnormal gait, dysphagia and muscle atrophy in affected horses [3].

The intracellular nature of S. neurona and its ability to evade the host’s immune surveillance [4] makes EPM treatment expensive, lengthy and challenging. Traditionally, clinical treatment of EPM
involved inhibitors of folate synthesis and metabolism (sulphonamides/pyrimethamine combination) over a prolonged period [5]. More recently, triazines derivatives (dicalazuril, ponazuril) that target the parasite’s apicoplast [6], nitazoxanide, a pyruvate:ferridoxin oxidoreductase analogue that inhibits the parasite’s anaerobic metabolism [7], and anti-inflammatory agents and immune stimulants [8] have been used with variable success in eliminating clinical signs. Despite the availability of these drugs, EPM treatment is complicated by the emergence of drug-resistance (due to intermittent or periodic treatments), cost of therapies and drug toxicity (e.g. diarrhea and/or anemia) and infection relapses due to re-growth of residual parasites after the treatment regimes [2]. As such, discovery and development of novel therapeutics for EPM is imperative.

To successfully invade the host cells, apicomplexans utilize three specialized exocytic organelles (micronemes, rhoptries and dense-granules) [9]. The microneme is used for host cell recognition, binding, penetration and gliding along the cytoskeletal structures. Rhoptry proteins are discharged into the host cell during parasite internalization and are crucial in the formation of the parasitophorous vacuoles (PVs). Developing zoites contain non-pedunculated condensing vesicles that synthesize and package inactive rhoptry proteins, which are proteolytically activated when the rhoptry contents are condensed [10]. The PVs facilitate parasite development by allowing nutrient transport from the host cell and by blocking lysosomal fusion, which would otherwise kill the parasites [11]. Upon internalization, zoites use the dense-granules to remodel the PVs into functionally active organelles.

The virulence of the apicomplexans is highly dependent on protein kinases (PKs) that are involved in the invasion and modification of host cell structure and function. PKs are classified into various groups, comprising of many families and sub-families [12-14]. The major apicomplexans PK groups are the PKs A, G and C (AGCs), which are regulated by cyclic nucleotides and other second messengers; the calmodulin/calcium-dependent PKs (CAMKs); the casein kinase 1 (CK1) and close relatives, the PKs involved in cell-cycle control and signal transduction (containing cyclin-dependent, mitogen-activated and cell division cycle [CDC]-like PKs; or CMGCs); the receptor guanylate cyclase (RGC); the “sterile” serine/threonine kinases (STEs, including several kinases functioning in the MAPK cascades); the tyrosine kinases (TKs); and the tyrosine kinase-like kinases (TKLs) [15-17]. In addition, there is a sizeable proportion of other PKs that do not fit into any of the above-mentioned eight major PK groups (i.e. the “Other” kinases [OPKs]), and another smaller set of atypical PKs (aPKs) that have no clear sequence similarity with the conventional PKs [13, 18, 19].

Each of the PK families have vital roles in parasite’s survival. For instance, PfPK-B (AGC family), PfTKL3 (TKL family), and four of the seven CDPKs (CAMK family) are required by Plasmodium parasites to complete their asexual cycle [20-22]. In a recent study, deletion of TgCK1α (CK1 family) resulted in defective replication of Toxoplasma gondii in vitro [23]. The P. falciparum CKL and SRPK1 (CMGC family) complement each other in the regulation of mRNA splicing (via phosphorylation of splicing factors [24]). Since apicomplexans lack typical MAPK cascades, the STEs are not well studied. However, in the parasites that do have the MAPK pathways, STEs are essential for parasite growth (e.g. in the human parasites, Schistosoma mansoni [25]). Further, for parasites without the conventional MAPK cascades, aPKs can activate the signaling pathways, for instance the Plasmodium Pfnek3 [26]. Finally, some of the notable OPKs include aurora kinases, rhoptry kinases (ROPKs) and parasite-specific eukaryotic initiation factor-2 (eIF2) kinases (eIF2K), which are important in parasite virulence and differentiation [27-29].

The differences between the parasite’s and their host’s PK homologs can be used to develop anti-parasitic strategies via specific inhibition of the parasite’s PKs [30]. This is feasible due to the essential roles of PKs in the parasite’s growth and development. Proof of principle of this approach has been demonstrated by the inhibition of human PKs using chemical ligands to treat cancers and other diseases [31, 32]. Recently, Ojo et al., [33] provided evidence that PKs can be targeted for rationally designed drugs that can potently inhibit growth of S. neurona. The technology is available and approved for therapeutic intervention, thus offering a unique prospect of repurposing chemical ligands to manage S. neurona infections [34].
Here, we used a genome-wide approach to define the kinome of *S. neurona* and determined the relatedness of the putative PKs to those reported in other apicomplexans. We also tested the docking of some specific inhibitors to a selection of some of the putative PKs. Defining the *S. neurona* kinome is not only important in providing insights into the parasite’s biology, but also identification of potential novel drug targets that can be used to clear chronic *S. neurona* infections, reduce parasite survival, augment the host’s anti-parasitic immune responses, or as candidates for development of EPM vaccines.

2. Results

2.1 *Sarcocystis neurona* encodes 92 putative kinases

To date, at least 15 apicomplexan genomes (coccidians, gregarines, hemosporidians and piroplasmids) have either been fully sequenced or partially annotated [35]. In the current study, we conducted an exhaustive genome-wide search of the newly sequenced *S. neurona* genome [36], and identified 92 putative PKs (Table 1). The identified PKs contained the characteristic PK (IPR000719) or PK-like (IPR011009) domains and three conserved amino acids constituting the catalytic triad (Lys30, Asp125, Asp143). The PKs had sizes ranging between 152 and 6544 amino acids and relative molecular weights of between 15.94 to 671.51 kDa. Majority of the PKs had isoelectric point (pI) greater than 7.0, implying that the PKs have low turnover rates since in general, acidic proteins are thought to be degraded more rapidly than neutral or basic proteins [37].

Assignment of *S. neurona* PK groups was accomplished by sequence clustering using Blast2GO [62], and by BLASTp searches in the Kinase.com database [14]. Out of the eleven known PK groups [12-14], *S. neurona* PKs segregated into the AGC (*n*=9), CAMK (*n*=20), CK1 (*n*=3), CMGC (*n*=19), STE (*n*=2), TKL (*n*=6), aPK (*n*=2) and OPK (*n*=31) (Table 1). Apart from the 31 OPKs that do not fit to the major kinase groups, the CAMK and CMGC groups, whose members are essential for the parasite’s host cell invasion [38] and differentiation (via cell-cycle regulation) [39], respectively, had the highest number of PKs underlying the importance of these processes in the parasite. Unlike in some parasites (e.g. *P. falciparum*) that lack STEs and TKLs [40], *S. neurona* contains PKs in the two classes.
Table 1. Description of the 92 putative PKs identified in the kinome of *S. neurona*. The putative PKs could be classified into eight groups. The amino acid coordinates of the conserved PK domains in the protein sequences, and the PK homologies to other apicomplexan PKs are shown in columns 7 to 12.

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Sequence annotations; description</th>
<th>Family; [subfamily]</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_1312</td>
<td>AGC kinase</td>
<td>AGC kinase</td>
<td>903</td>
<td>5.52</td>
<td>101.1</td>
<td>137–481</td>
<td>PDPK; [T. gondii RUB]</td>
<td>417</td>
<td>6.00E-13</td>
<td>59</td>
<td>KFG61374.1</td>
</tr>
<tr>
<td>SRCN_3339</td>
<td>PKA</td>
<td>3-phosphoinositide dependent PK-1 (PDK1)</td>
<td></td>
<td>5.52</td>
<td>101.1</td>
<td>137–481</td>
<td>Putative AGC kinase; [N. caninum L]</td>
<td>611</td>
<td>0.0</td>
<td>81</td>
<td>CEL65574.1</td>
</tr>
<tr>
<td>SRCN_4249</td>
<td>AGC kinase</td>
<td>Nuclear dbf2-related (NDR)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Putative AGC kinase; [N. caninum L]</td>
<td>219</td>
<td>2.00E-68</td>
<td>76</td>
<td>XP_003883757.1</td>
</tr>
<tr>
<td>SRCN_4518</td>
<td>PK G AGC kinase family member PKG</td>
<td>Ciliate-E2</td>
<td>425</td>
<td>5.83</td>
<td>48.78</td>
<td>97–399</td>
<td>AGC kinase TgPKG1; [T. gondii ME49]</td>
<td>830</td>
<td>0.0</td>
<td>92</td>
<td>EPR61116.1</td>
</tr>
<tr>
<td>SRCN_3990</td>
<td>cAMP-dependent kinase</td>
<td>CAMKL; [MELK]</td>
<td>1907</td>
<td>9.55</td>
<td>217.5</td>
<td>782–1634</td>
<td>cAMP-dependent protein kinase [T. gondii VEG]</td>
<td>75.9</td>
<td>1e-12</td>
<td>30</td>
<td>ESS31194.1</td>
</tr>
<tr>
<td>SRCN_5165</td>
<td>cAMP-dependent PK, catalytic chain</td>
<td>PKA</td>
<td>343</td>
<td>8.99</td>
<td>39.36</td>
<td>20–338</td>
<td>AGC kinase; [T. gondii ARI]</td>
<td>425</td>
<td>3.00E-15</td>
<td>90</td>
<td>KYF43224.1</td>
</tr>
<tr>
<td>SRCN_4913</td>
<td>Putative PK</td>
<td>PKD</td>
<td>2330</td>
<td>6.42</td>
<td>244.6</td>
<td>1093–1737</td>
<td>Putative PK; [E. tenella]</td>
<td>107</td>
<td>9.00E-22</td>
<td>60</td>
<td>XP_013228294.1</td>
</tr>
<tr>
<td>SRCN_5430</td>
<td>AGC kinase</td>
<td>Ribosomal protein S6 Kinases (RSK; [p70])</td>
<td></td>
<td>5.54</td>
<td>139.2</td>
<td>824–1344</td>
<td>AGC kinase; [T. gondii MAS]</td>
<td>223</td>
<td>3.00E-60</td>
<td>59</td>
<td>KFH07588.1</td>
</tr>
<tr>
<td>SRCN_5610</td>
<td>cAMP-dependent PK, catalytic chain</td>
<td>PKA</td>
<td>333</td>
<td>9.00</td>
<td>37.96</td>
<td>12–318</td>
<td>cAMP-dependent PK, catalytic subunit; [T. gondii ME49]</td>
<td>641</td>
<td>0.0</td>
<td>92</td>
<td>XP_002366464.1</td>
</tr>
</tbody>
</table>

2. Kinase Group calcium (Ca2+)/calmodulin-regulated kinases (CAMK)

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Sequence annotations; description</th>
<th>Family; [subfamily]</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_1071</td>
<td>Ca²⁺-dependent kinase</td>
<td>CAMK1</td>
<td>1495</td>
<td>9.47</td>
<td>152.6</td>
<td>1085–1401</td>
<td>Putative PK; [T. gondii VEG]</td>
<td>188</td>
<td>1.00E-47</td>
<td>67</td>
<td>ESS31884.1</td>
</tr>
<tr>
<td>SRCN_2032</td>
<td>Putative PK</td>
<td>Ciliate-C1</td>
<td>297</td>
<td>6.22</td>
<td>33.26</td>
<td>15–297</td>
<td>PK; [H. hammondii]</td>
<td>424</td>
<td>1.00E-14</td>
<td>68</td>
<td>XP_008882026.1</td>
</tr>
<tr>
<td>SRCN_2165</td>
<td>Ca²⁺-dependent kinase CDPK2B</td>
<td>CDPK</td>
<td>692</td>
<td>7.33</td>
<td>75.65</td>
<td>101–401</td>
<td>Ca²⁺-dependent PK CDPK2A; [T. gondii]</td>
<td>674</td>
<td>0.0</td>
<td>90</td>
<td>KFY44522.1</td>
</tr>
</tbody>
</table>
### Description of the putative protein kinases (PKs) in the genome of *S. neurona*

<table>
<thead>
<tr>
<th>Protein ID *</th>
<th>Sequence annotations; description *</th>
<th>Family; [subfamily] c</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_2257</td>
<td>Histone kinase</td>
<td>CAMKL; [AMP-activated protein kinase (AMPK)]</td>
<td>1800</td>
<td>8.98</td>
<td>187.7</td>
<td>1159–1448</td>
<td>Putative CAM kinase, SNF1 family; [E. acervulina]</td>
<td>377</td>
<td>7.00E-10</td>
<td>64</td>
<td>XP_013252246.1</td>
</tr>
<tr>
<td>SRCN_2544</td>
<td>CAM SNF1 AMK1 family</td>
<td>CAMKL; [AMPK-regulated kinase novel kinase (NUAK)]</td>
<td>333</td>
<td>5.71</td>
<td>37.85</td>
<td>62–333</td>
<td>CAM kinase, SNF1/AMK1 family ToxPK1; [N. caninum L]</td>
<td>480</td>
<td>9.00E-16</td>
<td>79</td>
<td>XP_003882065.1</td>
</tr>
<tr>
<td>SRCN_2937</td>
<td>Ca²⁺-signalling kinase MARK</td>
<td>CAMKL; [microtubule affinity regulating kinase (MARK)]</td>
<td>278</td>
<td>9.62</td>
<td>31.29</td>
<td>1–250</td>
<td>Putative Ca²⁺ signalling PK MARK; [T. gondii GT1]</td>
<td>356</td>
<td>5.00E-12</td>
<td>72</td>
<td>EPR59053.1</td>
</tr>
<tr>
<td>SRCN_3011</td>
<td>Calmodulin-dependent PK (CAM) CDPK6</td>
<td>CDPK</td>
<td>1435</td>
<td>9.25</td>
<td>154.4</td>
<td>1238–1435</td>
<td>Cdpk kinase domain; [T. gondii]</td>
<td>179</td>
<td>6.00E-49</td>
<td>75</td>
<td>3IS5_A</td>
</tr>
<tr>
<td>SRCN_3314</td>
<td>A Chain crystal</td>
<td>CDPK1</td>
<td>519</td>
<td>5.99</td>
<td>58.89</td>
<td>37–335</td>
<td>Calmodulin-domain PK 1; [T. gondii]</td>
<td>536</td>
<td>0.0</td>
<td>97</td>
<td>3MA6_A</td>
</tr>
<tr>
<td>SRCN_3583</td>
<td>Ca²⁺-dependent kinase CDPK5</td>
<td>CDPK</td>
<td>454</td>
<td>6.09</td>
<td>50.38</td>
<td>35–308</td>
<td>Ca²⁺-dependent PK CDPK5; [T. gondii ARI]</td>
<td>776</td>
<td>0.0</td>
<td>89</td>
<td>KYF43137.1</td>
</tr>
<tr>
<td>SRCN_3701</td>
<td>Ca²⁺-dependent kinase CDPK3</td>
<td>CDPK</td>
<td>560</td>
<td>5.91</td>
<td>62.07</td>
<td>77–362</td>
<td>Ca²⁺-dependent Kinase; [T. gondii]</td>
<td>493</td>
<td>1.00E-17</td>
<td>87</td>
<td>3DXN_A</td>
</tr>
<tr>
<td>SRCN_4076</td>
<td>CAM CDPK family</td>
<td>CDPK</td>
<td>1701</td>
<td>5.93</td>
<td>181.4</td>
<td>1079–1663</td>
<td>CAM kinase, CDPK family; [H. hammondii]</td>
<td>239</td>
<td>2.00E-62</td>
<td>69</td>
<td>XP_00888497.1</td>
</tr>
<tr>
<td>SRCN_4093</td>
<td>PK</td>
<td>CDPK</td>
<td>1155</td>
<td>9.02</td>
<td>118.3</td>
<td>1–261</td>
<td>Putative atypical MEK-related kinase; [N. caninum L]</td>
<td>222</td>
<td>1.00E-59</td>
<td>70</td>
<td>XP_00388069.1</td>
</tr>
<tr>
<td>SRCN_4390</td>
<td>Ca²⁺-dependent kinase CDPK2</td>
<td>CDPK</td>
<td>790</td>
<td>6.24</td>
<td>85.47</td>
<td>280–556</td>
<td>Ca²⁺-dependent PK, related; [N. caninum L] SNF1-related PK catalytic-α KIN10, 5 AMP-activated PK; [N. caninum L]</td>
<td>1130</td>
<td>0.0</td>
<td>74</td>
<td>XP_00388432.1</td>
</tr>
<tr>
<td>SRCN_4815</td>
<td>Histone kinase (partial)</td>
<td>CDPK</td>
<td>711</td>
<td>6.53</td>
<td>75.98</td>
<td>1–314</td>
<td>SNF1-related PK catalytic-α KIN10, 5 AMP-activated PK; [N. caninum L]</td>
<td>568</td>
<td>0.0</td>
<td>48</td>
<td>CEL67550.1</td>
</tr>
<tr>
<td>SRCN_2257</td>
<td>CAM CDPK</td>
<td>CDPK</td>
<td>2748</td>
<td>8.98</td>
<td>282.1</td>
<td>796–885</td>
<td>Putative CAM kinase,</td>
<td>117</td>
<td>3.00E-24</td>
<td>59</td>
<td>XP_00388190.1</td>
</tr>
</tbody>
</table>
### Description of the putative protein kinases (PKs) in the genome of *S. neurona*

<table>
<thead>
<tr>
<th>Protein ID *</th>
<th>Sequence annotations; description *</th>
<th>Family; [subfamily] c</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Description of protein kinase (PK) homologies (BLASTp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_5227</td>
<td>CDPK8-like</td>
<td></td>
<td>467</td>
<td>8.94</td>
<td>52.04</td>
<td>168–446</td>
<td>CDPK family; [N. caninum L]</td>
</tr>
<tr>
<td>SRCN_5410</td>
<td>Calmodulin-dependent PK (CAM-SNF1 family)</td>
<td>CAMK1</td>
<td>760</td>
<td>8.37</td>
<td>84.23</td>
<td>254–573</td>
<td>CAM kinase, SNF1 family; [H. hammondi]</td>
</tr>
<tr>
<td>SRCN_5812</td>
<td>Ca2+-dependent kinase CDPK9</td>
<td>CDPK</td>
<td>3298</td>
<td>7.11</td>
<td>345.8</td>
<td>208–860</td>
<td>Ca2+-dependent PK CDPK9; [H. hammondi]</td>
</tr>
<tr>
<td>SRCN_5948</td>
<td>Ca2+-dependent kinase CDPK8</td>
<td>CDPK</td>
<td>1374</td>
<td>9.09</td>
<td>138.2</td>
<td>365–623</td>
<td>PK-PH domain-containing protein; [T. gondii ME49]</td>
</tr>
<tr>
<td>SRCN_6597</td>
<td>Ca2+-dependent kinase CDPK7</td>
<td>CAMK1</td>
<td>1632</td>
<td>9.42</td>
<td>170.9</td>
<td>813–1236</td>
<td>Ca2+-dependent PK; [T. gondii]</td>
</tr>
<tr>
<td>SRCN_6606</td>
<td>Ca2+-dependent kinase CDPK4</td>
<td>CDPK</td>
<td>372</td>
<td>9.34</td>
<td>37.65</td>
<td>6–290</td>
<td>Casein kinase 1; [T. gondii ME49]</td>
</tr>
<tr>
<td>SRCN_4587</td>
<td>Casein kinase I</td>
<td>CK1-D</td>
<td>137</td>
<td>7.78</td>
<td>15.94</td>
<td>1–137</td>
<td>Casein kinase I; [H. hammondi]</td>
</tr>
<tr>
<td>SRCN_4645</td>
<td>Casein kinase I</td>
<td>CK1-D</td>
<td>229</td>
<td>9.51</td>
<td>25.66</td>
<td>38–229</td>
<td>Casein kinase I; [T. gondii GAB2-2007-GAL-DOM 2]</td>
</tr>
<tr>
<td>SRCN_1104</td>
<td>Cyclin-dependent kinase family 5 (CDKL)</td>
<td>CK1-D</td>
<td>372</td>
<td>9.23</td>
<td>42.71</td>
<td>1–318</td>
<td>Cyclin-dependent kinase family 5 protein; [H. hammondi]</td>
</tr>
<tr>
<td>SRCN</td>
<td>Cell-cycle-associated Serine-arginine rich PK</td>
<td></td>
<td>2911</td>
<td>5.37</td>
<td>302.9</td>
<td>713–1837</td>
<td>PK; [T. gondii ME49]</td>
</tr>
</tbody>
</table>

### Kinase Group casein kinase 1 (cell kinase 1)

| SRCN_3445  | Casein kinase I                    | CK1-D               | 323         | 9.34 | 37.65   | 6–290                 | Casein kinase 1; [T. gondii ME49]                 |
| SRCN_4587  | Casein kinase I                    | CK1-D               | 137         | 7.78 | 15.94   | 1–137                 | Casein kinase I; [H. hammondi]                   |
| SRCN_4645  | Casein kinase I                    | CK1-D               | 229         | 9.51 | 25.66   | 38–229                | Casein kinase I; [T. gondii GAB2-2007-GAL-DOM 2] |

### Kinase Group CMGC (including cyclin-dependent kinases, mitogen-activated PKs, glycogen synthase kinases and CDK-like kinases)

| SRCN_1104  | Cyclin-dependent kinase family 5 (CDKL) | CK1-D               | 372         | 9.23 | 42.71   | 1–318                 | Cyclin-dependent kinase family 5 protein; [H. hammondi] |
| SRCN        | Cell-cycle-associated Serine-arginine rich PK |                  | 2911        | 5.37 | 302.9   | 713–1837              | PK; [T. gondii ME49]                              |

*Protein ID*: The identification number for each protein.

*Sequence annotations; description*: Additional information about the protein sequence.

*Family; [subfamily]*: The family and subfamily classification of the protein.

*Length [aa]*: The length of the protein in amino acids.

*pI*: The isoelectric point of the protein.

*MW [kDa]*: The molecular weight of the protein in kiloDaltons.

*PK domain coordinates*: The coordinates of the PK domain.

*Sequence name; [apicomplexan]*: The name of the sequence and its source.

*Bit Score*: The bit score of the BLASTp alignment.

*E-value*: The E-value of the BLASTp alignment.

*Identity [%]*: The percentage identity of the aligned sequences.

*Accession number*: The accession number of the sequence in the database.
### Description of the putative protein kinases (PKs) in the genome of *S. neurona*

| Protein ID | Sequence annotations; description | Family; [subfamily] | Length [aa] | pI | MW [kDa] | PK domain coordinates | Sequence name; [apicomplexan] | Bit Score | E-value | Identity [%] | Accession number |
|------------|---------------------------------|---------------------|-------------|----|---------|-----------------------|-------------------------------|----------|---------|--------------|----------------|-----------------|
| SRCN_1236  | kinase (SRPK)                   | (SRPK)              |             |    |         |                       | Cell-cycle-associated PK CLK; [T. gondii FOU] | 288      | 2.00E-81 | 74           | KFG33061.1      |
| SRCN_1479  | CMGC Lammer kinase              | CLK                 | 748         | 10.2| 79.09   | 485–748               | cell-cycle-associated PK CLK; [T. gondii FOU] | 223      | 2.00E-57 | 63           | ESS33160.1      |
| SRCN_1611  | CMGC Dual-specificity tyrosine-regulated kinase (Dyrk) | Dyrk; [DyrkP] | 1504         | 5.86| 160.3   | 551–1498              | Cell-cycle-associated PK Dyrk; [T. gondii VEG] | 223      | 2.00E-57 | 63           | ESS33160.1      |
| SRCN_1731  | Cell-cycle-associated kinase GSK | Glycogen synthase kinase (GSK) | 219         | 6.59| 24.39   | 1–175                 | Cell-cycle-associated PK GSK; [H. hammondii] | 330      | 2.00E-11 | 82           | XP_008887193.1  |
| SRCN_1732  | Cell-cycle-associated kinase GSK | Glycogen synthase kinase (GSK) | 203         | 10.7| 20.72   | 86–203                | CMGC kinase, GSK family TgPK3; [E. brunetti] | 114      | 1.00E-28 | 91           | CDJ46527.1      |
| SRCN_2759  | Cell-cycle-associated kinase partial | Ca²⁺-dependent PK (CDK); [CRK7] | 1122        | 6.14| 118.6   | 7                     | Cell-cycle-associated PK CDK; [T. gondii VAND] | 118      | 1.00E-25 | 79           | KFH12036.1      |
| SRCN_2845  | CMGC DYRK PRP4 kinase           | DYRK; [PRP4]        | 1665        | 9.76| 177.0   | 8                     | Putative PK (CLK3); [P. malaia] | 330      | 1.00E-10 | 69           | SBS85334.1      |
| SRCN_3891  | CMGC kinase                     | DYRK; [DYRK2]       | 674         | 8.85| 73.85   | 399–764               | Putative CMGC kinase; [T. gondii ME49] | 80.1     | 6.00E-14 | 67           | EPT25192.1      |
| SRCN_4209  | CMGC MAPK family (ERK) MAPK-1   | Mitogen-activated PK (MAPK); [ERK]) | 2361        | 6.73| 247.7   | 94–754                | CMGC, MAPK/ (ERK) TgMAPK-1; [E. brunetti] | 137      | 1.00E-30 | 74           | CDJ49492.1      |
| SRCN_4674  | Cyclin-dependent kinase         | Ca²⁺-dependent PK (CDK); [CDK7] | 138         | 7.80| 15.50   | 1–138                 | Cyclin-dependent kinase; [T. gondii GT1] | 108      | 7.00E-27 | 58           | EPR60430.1      |
| SRCN_4801  | Cell-cycle-associated kinase    | Ca²⁺-dependent PK (CDK); [CDK5] | 300         | 6.08| 34.33   | 1–289                 | CMGC kinase, CDK family TgPK2; [N. caninum L] | 576      | 0.0      | 91           | XP_003885801.1  |
| SRCN_5365  | Cell-cycle-associated kinase MAPK | Mitogen-activated PK (MAPK; [ERK]) | 417         | 6.77| 48.32   | 7–363                 | Cell-cycle-associated PK MAPK; [H. hammondii] | 823      | 0.0      | 93           | XP_008886907.1  |
| SRCN_5635  | Cell-cycle-associated kinase MAPK | Mitogen-activated PK (MAPK; [ERK]) | 690         | 9.55| 80.90   | 208–603               | Putative              | 390      | 2.00E-12 | 87           | KFY45878.1      |

*Bit Score* - Bit score of the alignment.
*E-value* - E-value of the alignment.
*Identity [%]* - Identity percentage of the alignment.
*Accession number* - Accession number of the protein.
<table>
<thead>
<tr>
<th>Protein ID (^a)</th>
<th>Sequence annotations; description (^b)</th>
<th>Family; [subfamily](^c)</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>6346</em> (\text{kinase CDK})</td>
<td>(CDK); [CDK5]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRCN _6427</td>
<td>CMGC CK2 kinase</td>
<td>Cell Kinase 2 (CK2)</td>
<td>1395</td>
<td>10.2</td>
<td>144.8</td>
<td>885–1356</td>
<td>cell-cycle-associated PK CDK; [T. gondii ARI]</td>
<td>241</td>
<td>6.00E-73</td>
<td>98</td>
<td>KFH07655.1</td>
</tr>
<tr>
<td>SRCN _6472</td>
<td>Cell-cycle-associated kinase ERK7</td>
<td>Mitogen-activated PK (MAPK; [ERK])</td>
<td>983</td>
<td>9.28</td>
<td>104.9</td>
<td>7–317</td>
<td>Cell-cycle-associated PK ERK7; [T. gondii ARI]</td>
<td>647</td>
<td>0.0</td>
<td>81</td>
<td>KFY46268.1</td>
</tr>
<tr>
<td>SRCN _761</td>
<td>Cell-cycle-associated kinase</td>
<td>Ca(^{2+})-dependent PK (CDK); [CDK7]</td>
<td>577</td>
<td>9.34</td>
<td>58.39</td>
<td>144–490</td>
<td>Cell-cycle-associated PK; [H. hammondii]</td>
<td>283</td>
<td>3.00E-88</td>
<td>68</td>
<td>XP_008882409.1</td>
</tr>
<tr>
<td>SRCN _895</td>
<td>Cell-cycle-associated kinase</td>
<td>Ca(^{2+})-dependent PK (CDK); [CDK10]</td>
<td>340</td>
<td>8.93</td>
<td>38.57</td>
<td>1–307</td>
<td>Cell-cycle-associated PK; [T. gondii ARI]</td>
<td>234</td>
<td>6.00E-75</td>
<td>76</td>
<td>KFY44017.1</td>
</tr>
<tr>
<td>SRCN _977</td>
<td>Cell-cycle-associated kinase CDK</td>
<td>Ca(^{2+})-dependent PK (CDK); [PITSLRE/CDK11]</td>
<td>1502</td>
<td>7.38</td>
<td>156.3</td>
<td>1114–1429</td>
<td>Cell-cycle-associated PK CDK; [T. gondii p89]</td>
<td>454</td>
<td>3.00E-13</td>
<td>92</td>
<td>KFG28420.1</td>
</tr>
<tr>
<td>SRCN _108</td>
<td>Unc-51-like autophagy activating kinase 1 (ULK1)</td>
<td>ULK</td>
<td>343</td>
<td>7.13</td>
<td>38.90</td>
<td>1-223</td>
<td>ULK kinase; [T. gondii VAND]</td>
<td>376</td>
<td>2.00E-13</td>
<td>75</td>
<td>KFH07419.1</td>
</tr>
<tr>
<td>SRCN _1606</td>
<td>eIF2 kinase IF2K-C</td>
<td>PEK; [general control nonderepressible 2 (GCN2)]</td>
<td>4034</td>
<td>8.98</td>
<td>406.5</td>
<td>1235–2178</td>
<td>eIF2 kinase IF2K-C; [T. gondii ARI]</td>
<td>259</td>
<td>4.00E-67</td>
<td>35</td>
<td>KFH07289.1</td>
</tr>
<tr>
<td>SRCN _2076</td>
<td>Rhopty kinase family ROP30</td>
<td>Conserved hypothetical protein</td>
<td>1276</td>
<td>9.18</td>
<td>134.7</td>
<td>812–1260</td>
<td>ROP30; [T. gondii VEG]</td>
<td>230</td>
<td>3.00E-63</td>
<td>53</td>
<td>CEL76436.1</td>
</tr>
<tr>
<td>SRCN _2123</td>
<td>Rhopty kinase family ROP35</td>
<td>PLK; [PLK-Unclassified]</td>
<td>291</td>
<td>9.30</td>
<td>33.50</td>
<td>53–265</td>
<td>ROP35; [T. gondii RUB]</td>
<td>207</td>
<td>2.00E-61</td>
<td>43</td>
<td>KFG59037.1</td>
</tr>
<tr>
<td>SRCN _3216</td>
<td>Rhopty kinase family ROP32</td>
<td>CAMK-Unique</td>
<td>523</td>
<td>7.08</td>
<td>57.00</td>
<td>214–520</td>
<td>Putative PK; [T. gondii VAND]</td>
<td>167</td>
<td>1.00E-42</td>
<td>30</td>
<td>KFH00232.1</td>
</tr>
<tr>
<td>SRCN _2183</td>
<td>Rhopty kinase family ROP35</td>
<td>Aurorla-like</td>
<td>226</td>
<td>6.36</td>
<td>25.84</td>
<td>1–212</td>
<td>ROP35; [T. gondii VEG]</td>
<td>198</td>
<td>8.00E-59</td>
<td>48</td>
<td>ESS33297.1</td>
</tr>
<tr>
<td>SRCN _2271</td>
<td>Putative PK (incomplete catalytic)</td>
<td>NimA (Never in mitosis gene A)-related Kinase</td>
<td>1463</td>
<td>9.02</td>
<td>157.1</td>
<td>437–1145</td>
<td>Putative PK; [N. caninum L]</td>
<td>327</td>
<td>5.00E-90</td>
<td>68</td>
<td>XP_003881849.1</td>
</tr>
</tbody>
</table>

### 5. Kinase Group ‘Other’ (OPK; i.e. kinases with conventional PK [ePK] domains that do not fit into any of the other major groups of kinases)

<table>
<thead>
<tr>
<th>Protein ID (^a)</th>
<th>Sequence annotations; description (^b)</th>
<th>Family; [subfamily]</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN _108</td>
<td>Unc-51-like autophagy activating kinase 1 (ULK1)</td>
<td>ULK</td>
<td>343</td>
<td>7.13</td>
<td>38.90</td>
<td>1-223</td>
<td>ULK kinase; [T. gondii VAND]</td>
<td>376</td>
<td>2.00E-13</td>
<td>75</td>
<td>KFH07419.1</td>
</tr>
<tr>
<td>SRCN _1606</td>
<td>eIF2 kinase IF2K-C</td>
<td>PEK; [general control nonderepressible 2 (GCN2)]</td>
<td>4034</td>
<td>8.98</td>
<td>406.5</td>
<td>1235–2178</td>
<td>eIF2 kinase IF2K-C; [T. gondii ARI]</td>
<td>259</td>
<td>4.00E-67</td>
<td>35</td>
<td>KFH07289.1</td>
</tr>
<tr>
<td>SRCN _2076</td>
<td>Rhopty kinase family ROP30</td>
<td>Conserved hypothetical protein</td>
<td>1276</td>
<td>9.18</td>
<td>134.7</td>
<td>812–1260</td>
<td>ROP30; [T. gondii VEG]</td>
<td>230</td>
<td>3.00E-63</td>
<td>53</td>
<td>CEL76436.1</td>
</tr>
<tr>
<td>SRCN _2123</td>
<td>Rhopty kinase family ROP35</td>
<td>PLK; [PLK-Unclassified]</td>
<td>291</td>
<td>9.30</td>
<td>33.50</td>
<td>53–265</td>
<td>ROP35; [T. gondii RUB]</td>
<td>207</td>
<td>2.00E-61</td>
<td>43</td>
<td>KFG59037.1</td>
</tr>
<tr>
<td>SRCN _3216</td>
<td>Rhopty kinase family ROP32</td>
<td>CAMK-Unique</td>
<td>523</td>
<td>7.08</td>
<td>57.00</td>
<td>214–520</td>
<td>Putative PK; [T. gondii VAND]</td>
<td>167</td>
<td>1.00E-42</td>
<td>30</td>
<td>KFH00232.1</td>
</tr>
<tr>
<td>SRCN _2183</td>
<td>Rhopty kinase family ROP35</td>
<td>Aurorla-like</td>
<td>226</td>
<td>6.36</td>
<td>25.84</td>
<td>1–212</td>
<td>ROP35; [T. gondii VEG]</td>
<td>198</td>
<td>8.00E-59</td>
<td>48</td>
<td>ESS33297.1</td>
</tr>
<tr>
<td>SRCN _2271</td>
<td>Putative PK (incomplete catalytic)</td>
<td>NimA (Never in mitosis gene A)-related Kinase</td>
<td>1463</td>
<td>9.02</td>
<td>157.1</td>
<td>437–1145</td>
<td>Putative PK; [N. caninum L]</td>
<td>327</td>
<td>5.00E-90</td>
<td>68</td>
<td>XP_003881849.1</td>
</tr>
</tbody>
</table>
## Description of the putative protein kinases (PKs) in the genome of *S. neurona*

<table>
<thead>
<tr>
<th>Protein ID *</th>
<th>Sequence annotations; description *</th>
<th>Family; [subfamily]</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_2403</td>
<td>Aurora kinase (incomplete catalytic triad) (NEK)</td>
<td>PLK; [SAK/Plk4]</td>
<td>778</td>
<td>9.79</td>
<td>79.92</td>
<td>492–778</td>
<td>Putative Aurora kinase; [N. caninum L]</td>
<td>127</td>
<td>3.00E-28</td>
<td>44</td>
<td>XP_003880644.1</td>
</tr>
<tr>
<td>SRCN_2630</td>
<td>NimA related kinase (NEK) family protein</td>
<td>NEK</td>
<td>351</td>
<td>8.70</td>
<td>38.38</td>
<td>1–336</td>
<td>NEK kinase; [T. gondii ME49]</td>
<td>242</td>
<td>8.00E-75</td>
<td>52</td>
<td>XP_018638958.1</td>
</tr>
<tr>
<td>SRCN_286</td>
<td>Wee kinase</td>
<td>Inhibitory regulator of the RAS-cAMP (IRA1) kinase suppressor (IKS)</td>
<td>1019</td>
<td>6.20</td>
<td>106.6</td>
<td>598–959</td>
<td>Wee kinase; [H. hammondi]</td>
<td>445</td>
<td>5.00E-14</td>
<td>58</td>
<td>XP_008882669.1</td>
</tr>
<tr>
<td>SRCN_3075</td>
<td>Tyrosine kinase-like (TKL) protein</td>
<td>Numb-associated kinase (NAK)</td>
<td>1571</td>
<td>8.41</td>
<td>164.1</td>
<td>16–500</td>
<td>TKL; [T. gondii TgCatPRC2]</td>
<td>138</td>
<td>1.00E-32</td>
<td>73</td>
<td>KYK64203.1</td>
</tr>
<tr>
<td>SRCN_3142</td>
<td>PIK3R4 kinase-related</td>
<td>Aurora</td>
<td>997</td>
<td>8.72</td>
<td>106.5</td>
<td>548–899</td>
<td>Putative PIK3R4 kinase-related protein; [N. caninum L]</td>
<td>449</td>
<td>2.00E-13</td>
<td>60</td>
<td>XP_003885774.1</td>
</tr>
<tr>
<td>SRCN_3151</td>
<td>NimA related kinase (NEK) family protein</td>
<td>NEK</td>
<td>3186</td>
<td>7.96</td>
<td>318.6</td>
<td>352–656</td>
<td>NEK kinase; [T. gondii VEG]</td>
<td>468</td>
<td>7.00E-13</td>
<td>73</td>
<td>CEL78174.1</td>
</tr>
<tr>
<td>SRCN_3247</td>
<td>Rhoptry kinase family</td>
<td>Ciliate-D</td>
<td>345</td>
<td>8.94</td>
<td>38.81</td>
<td>23–325</td>
<td>ROP27; [T. gondii p89]</td>
<td>163</td>
<td>4.00E-43</td>
<td>31</td>
<td>KFG37427.1</td>
</tr>
<tr>
<td>SRCN_3417</td>
<td>Aurora kinase</td>
<td>Aurora</td>
<td>438</td>
<td>7.65</td>
<td>48.49</td>
<td>14–289</td>
<td>Aurora kinase; [T. gondii TgCatPRC2]</td>
<td>490</td>
<td>3.00E-15</td>
<td>76</td>
<td>KYK63669.1</td>
</tr>
<tr>
<td>SRCN_3444</td>
<td>Unc-51-like Autophagy activating kinase 1 (ULK1)</td>
<td>ULK</td>
<td>406</td>
<td>6.52</td>
<td>44.69</td>
<td>12–406</td>
<td>ULK kinase; [T. gondii RUB]</td>
<td>232</td>
<td>3.00E-71</td>
<td>61</td>
<td>KFG59767.1</td>
</tr>
<tr>
<td>SRCN_3669</td>
<td>CMGC kinase</td>
<td>ULK</td>
<td>1803</td>
<td>8.41</td>
<td>189.6</td>
<td>736–1200</td>
<td>Putative CMGC kinase; [N. caninum L]</td>
<td>624</td>
<td>0.0</td>
<td>62</td>
<td>CEL65030.1</td>
</tr>
<tr>
<td>SRCN_4410</td>
<td>Rhoptry kinase family</td>
<td>PKA-like</td>
<td>204</td>
<td>9.44</td>
<td>23.50</td>
<td>1–166</td>
<td>ROP35; [H. hammondi]</td>
<td>107</td>
<td>1.00E-25</td>
<td>39</td>
<td>XP_008885989.1</td>
</tr>
<tr>
<td>SRCN_4503</td>
<td>eIF2 kinase IF2K-B</td>
<td>PEK; [general control nonderepressible 2 (GCN2)]</td>
<td>158</td>
<td>5.76</td>
<td>17.59</td>
<td>1–158</td>
<td>eIF2 kinase IF2K-B [T. gondii TgCatPRC2]</td>
<td>149</td>
<td>5.00E-41</td>
<td>74</td>
<td>KYK69938.1</td>
</tr>
<tr>
<td>SRCN_4500</td>
<td>NimA related kinase</td>
<td>NEK</td>
<td>187</td>
<td>8.20</td>
<td>21.23</td>
<td>1–186</td>
<td>NEK kinase;</td>
<td>177</td>
<td>4.00E-54</td>
<td>64</td>
<td>XP_008885186.1</td>
</tr>
</tbody>
</table>
### Description of the putative protein kinases (PKs) in the genome of *S. neurona*

<table>
<thead>
<tr>
<th>Protein ID *</th>
<th>Sequence annotations; description b</th>
<th>Family; [subfamily] c</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_4528</td>
<td>(NEK) family protein Aurora kinase (incomplete catalytic triad)</td>
<td>Serum and glucocorticoid induced Kinase (SGK)</td>
<td>295</td>
<td>8.81</td>
<td>31.42</td>
<td>1–249</td>
<td>Putative Aurora kinase; [N. caninum L]</td>
<td>126</td>
<td>7.00E-31</td>
<td>43</td>
<td>CEL65223.1</td>
</tr>
<tr>
<td>SRCN_2404</td>
<td>SRCN_5653 NIMA-related kinase NIMA1</td>
<td>Aurora</td>
<td>626</td>
<td>8.27</td>
<td>60.75</td>
<td>513–626</td>
<td>PEK kinase [T. gondii TgCatPRC2]</td>
<td>251</td>
<td>1.00E-76</td>
<td>60</td>
<td>KYK62422.1</td>
</tr>
<tr>
<td>SRCN_5943</td>
<td>SRCN_6157 Unc-51-like autophagy activating kinase 1 (ULK1)</td>
<td>ULK</td>
<td>2420</td>
<td>9.38</td>
<td>250.5</td>
<td>8 1380–1672</td>
<td>ULK kinase [T. gondii ME49]</td>
<td>99.8</td>
<td>3e-21</td>
<td>38</td>
<td>XP_018635814.1</td>
</tr>
<tr>
<td>SRCN_6184</td>
<td>SRCN_6572 PK</td>
<td>ULK</td>
<td>199</td>
<td>6.74</td>
<td>22.72</td>
<td>1–183</td>
<td>PK; [H. hammondi]</td>
<td>172</td>
<td>3.00E-48</td>
<td>53</td>
<td>XP_00888491.1</td>
</tr>
<tr>
<td>SRCN_4310</td>
<td>SRCN_7082 Rhooptrry kinase family ROP33</td>
<td>Kinase Homologous to SPS1/STE20 (KHS)</td>
<td>1591</td>
<td>9.85</td>
<td>169.6</td>
<td>3 1265–1578</td>
<td>ROP33; [H. hammondi]</td>
<td>306</td>
<td>2.00E-87</td>
<td>39</td>
<td>XP_008887632.1</td>
</tr>
<tr>
<td>SRCN_7084</td>
<td>SRCN_7084 Rhooptrry kinase family ROP37</td>
<td>Ribosomal protein S6 Kinases (RSK; [RSK])</td>
<td>339</td>
<td>5.41</td>
<td>38.09</td>
<td>19–334</td>
<td>ROP37; [N. caninum L]</td>
<td>144</td>
<td>1.00E-36</td>
<td>36</td>
<td>CEL64242.1</td>
</tr>
</tbody>
</table>

6. Kinase Group "Sterile" serine/threonine kinase, or sterile-phenotype kinases (STE)

<table>
<thead>
<tr>
<th>Protein ID *</th>
<th>Sequence annotations; description b</th>
<th>Family; [subfamily] c</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_1328</td>
<td>Serine threonine kinase &quot;Sterile&quot;</td>
<td>Conserved hypothetical protein</td>
<td>1461</td>
<td>9.29</td>
<td>158.7</td>
<td>4 559–671</td>
<td>Hypothetical protein, conserved; [E. maxima]</td>
<td>88.6</td>
<td>5.00E-16</td>
<td>68</td>
<td>XP_013335801.1</td>
</tr>
<tr>
<td>SRCN_5172</td>
<td>SRCN_5172 serine/threonine 20-like (MST)</td>
<td>Mammalian Sterile</td>
<td>6552</td>
<td>6.14</td>
<td>671.5</td>
<td>1 3410–4122</td>
<td>STE kinase; [T. gondii TgCatPRC2]</td>
<td>412</td>
<td>1.00E-11</td>
<td>54</td>
<td>KYK71951.1</td>
</tr>
</tbody>
</table>

---

* * *
<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Sequence annotations; description</th>
<th>Family; [subfamily] c</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_1435</td>
<td>Tyrosine kinase-like (TKL)</td>
<td>Mixed lineage kinase (MLK); [Leucine Zipper-bearing Kinase (LZK)]</td>
<td>3064</td>
<td>8.05</td>
<td>306.7</td>
<td>2540–3060</td>
<td>Tyrosine kinase-like (TKL) protein; [N. caninum L]</td>
<td>278</td>
<td>3.00E-73</td>
<td>72</td>
<td>CEL64955.1</td>
</tr>
<tr>
<td>SRCN_1571</td>
<td>Tyrosine kinase-like (TKL)</td>
<td>Microtubule-associated S/T kinase (MAST)</td>
<td>550</td>
<td>9.76</td>
<td>59.91</td>
<td>135–501</td>
<td>Conserved hypothetical protein; [E. praecox]</td>
<td>76.3</td>
<td>6.00E-13</td>
<td>53</td>
<td>CDI87140.1</td>
</tr>
<tr>
<td>SRCN_3466</td>
<td>Tyrosine kinase-like (TKL)</td>
<td>TKL-Unique</td>
<td>3002</td>
<td>9.87</td>
<td>320.9</td>
<td>2342–2997</td>
<td>Tyrosine kinase-like (TKL) protein; [H. hammondi]</td>
<td>202</td>
<td>3.00E-50</td>
<td>65</td>
<td>XP_008887506.1</td>
</tr>
<tr>
<td>SRCN_3928</td>
<td>Tyrosine kinase-like (TKL)</td>
<td>LISK - LIMK (LIM kinase) and TESK (Testicular protein Kinase); [DD1]</td>
<td>5842</td>
<td>8.78</td>
<td>608.8</td>
<td>3639–4268</td>
<td>Tyrosine kinase-like (TKL) protein; [T. gondii TgCatPRC2]</td>
<td>216</td>
<td>2.00E-61</td>
<td>79</td>
<td>KYK63216.1</td>
</tr>
<tr>
<td>SRCN_4277</td>
<td>Kinase domain-containing protein</td>
<td>TKL-ciliate1</td>
<td>2256</td>
<td>8.43</td>
<td>240.2</td>
<td>1570–2256</td>
<td>Tyrosine kinase-like (TKL) protein; [N. caninum L]</td>
<td>204</td>
<td>8.00E-51</td>
<td>61</td>
<td>CEL67693.1</td>
</tr>
<tr>
<td>SRCN_811</td>
<td>Tyrosine kinase-like (TKL)</td>
<td>TKL-Unique</td>
<td>1099</td>
<td>9.18</td>
<td>119.2</td>
<td>814–1083</td>
<td>Putative tyrosine kinase-like (TKL) protein; [E. acervulina]</td>
<td>403</td>
<td>6.00E-12</td>
<td>59</td>
<td>XP_013252162.1</td>
</tr>
</tbody>
</table>

8. Kinase Group Atypical (aPKs)

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Sequence annotations; description</th>
<th>Family; [subfamily] c</th>
<th>Length [aa]</th>
<th>pI</th>
<th>MW [kDa]</th>
<th>PK domain coordinates</th>
<th>Sequence name; [apicomplexan]</th>
<th>Bit Score</th>
<th>E-value</th>
<th>Identity [%]</th>
<th>Accession number</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_3601</td>
<td>Atypical MEK-related kinase TRIO</td>
<td>Muscle-associated kinase</td>
<td>950</td>
<td>7.20</td>
<td>103.1</td>
<td>381–850</td>
<td>Atypical MEK-related kinase; [T. gondii GT1]</td>
<td>171</td>
<td>1.00E-42</td>
<td>32</td>
<td>EPR62774.1</td>
</tr>
<tr>
<td>SRCN_5962</td>
<td>Atypical MEK-related kinase ROCK-like</td>
<td>Rho-associated protein kinase (ROCK-like)</td>
<td>805</td>
<td>5.01</td>
<td>87.71</td>
<td>525–805</td>
<td>Atypical MEK-related kinase; [H. hammondi]</td>
<td>127</td>
<td>4.00E-29</td>
<td>56</td>
<td>XP_008884362.1</td>
</tr>
</tbody>
</table>

a The protein sequences and their corresponding identified were obtained from Toxoplasma Genomics Resource database (Release 28; version May 2016) [41]; b The descriptions of the protein sequence is based on BLASTp annotations using Blast2GO (see text for details); c The kinase classification is based on BLASTp on the kinase database.
2.1.1 The AGC group

The numbers of apicomplexan AGCs range from four (in *B. bovis*) to 15 (in *T. gondii*) [15]. Based on our Blast2GO annotations and BLASTp homology searches against the kinome database, five out of the nine *S. neurona* AGCs (SRCN_3339, SRCN_3990, SRCN_5165, SRCN_5610 and SRCN_1312) were homologs to the universally conserved PKAs that are found in *N. caninum* and *T. gondii* (see Table 1). The PKAs are essential for completion of schizogony (asexual reproduction) in *Plasmodium* parasites [42]. Further, *S. neurona* contains a putative PKG (SRCN_4518) which shows high homology (92%) to the *T. gondii* TgPKG1 (Table 1); PKGs are in essential apicomplexans [43].

2.1.2 The CAMK group

CAMKs form the second-largest apicomplexan PKs (after OPKs). Apicomplexan kinomes constitute varying numbers of CAMKs, which range from seven (in *B. bovis*) to 29 (in *T. gondii*) [15]. The most important CAMK family is the CDPK, which appeared to constitute almost 50% of *S. neurona* putative CAMKs (See Table 1). In terms of homologies, the *S. neurona* kinome contained orthologs to the *T. gondii* CDPK1 (SCRN_3314), CDPK2B (SCRN_2165), CDPK3 (SCRN_3701), CDPK4 (SCRN_6606), CDPK5 (SCRN_3583), CDPK6 (SCRN_3011), CDPK7 (SCRN_6597) and CDPK8 (SCRN_5948). Other CDPK orthologs were to the *N. caninum* CDPK2 (SCRN_4390) and *H. hammondi* CDPK9 (SCRN_5812) (Table 1). Inhibition of TgCDPK1 has been shown to disrupt the motility, host cell invasion and egress of *T. gondii* [44]. Owing to the absence of mammalian CDPK homologs, the identification of a relatively large number of CDPK homologs in *S. neurona* could be utilized in the rational design of anti-parasitic therapeutics.

2.1.3 The CK1 group

It is notable that *S. neurona* putatively encodes for three CK1 enzymes. Apart from *T. gondii* and some alveolates (e.g. *Cryptosporidium hominis* and *C. parvum*, important causative agents of diarrhea in children), which have three and two CK1 enzymes, respectively, most apicomplexans possess a single CK1 enzyme [15]. Two of the three *S. neurona* putative CK1 (SRCN_3445 and SRCN_4645) showed high sequence similarity (>90%) to the *T. gondii* TME49_040640 (TgCK1-α) and TGME49_089320 (TgCK1-β), respectively (Table 1). Inhibition of CK1 showed potential for anti-parasitic interventions in *T. gondii* [45]. CK1 is critical for the asexual proliferation of the *Plasmodium* parasites and is expressed in all the life-cycle stages of the parasite [46]. Three putative *S. neurona* CK1 had significant sequence similarity to the *P. falciparum* PfCK1, i.e. 74% (SRCN_3445), 65% (SRCN_4587) and 56% (SRCN_4645) (data not shown).

2.1.4 The CMGC group

The CMGC is the largest PK group in apicomplexans; CMGC numbers range from 15 in *B. bovis* to 23 in *P. vivax* [15], which is within the range we identified in the *S. neurona* kinome in our study (i.e. 19 CMGCs; See Table 1). Notable of these were the two GSK homologs (SRCN_1731 and SRCN_1732). This finding is similar to what has been observed in *Plasmodium* parasites in which two GSK-3 enzymes have been reported, both of which are essential for the parasite [47]. Homology searches showed considerable sequence similarity (51% and 41% for SRCN_1731 and SRCN_1732, respectively) to the PfGSK-3 enzymes (data not shown). Notably, eight of the 19 CMGCs in *S. neurona* were CDKs, including CDK7 (SRCN_4674, SRCN_2759 and SRCN_761), CDK10 (SRCN_895) and CDK11 (SRCN_977). Available data show that CDKs are essential in *P. falciparum* [35]. We also identified two putative MAPK homologs (SRCN_4209 and SRCN_5365), and ERK7 (SRCN_6472) (see Table 1), a result which is comparable to the two MAPKs in the kinome of *P. falciparum* [15].
2.1.5 The OPK group

The apicomplexan-specific OPKs are a tight cluster of PKs without clear relation to any of the other major PK groups. Notable of these are ROPKs which have high sequence divergence and have been thought to be largely restricted to *T. gondii* [48], which has a total of 34 members spread in over 40 distinct sub-families [27]. Although their diversification in apicomplexans is poorly understood, some ROPKs are key virulence factors in *T. gondii* [27]. At least nine putative ROPKs could be identified in *S. neurona*, including ROPK19A (SRCN_6184), ROP27 (SRCN_3247), ROP30 (SRCN_2076), ROP33 (SRCN_7082 and SRCN_7086), ROP35 (SRCN_2183, SRCN_2123, SRCN_7083 and SRCN_4410) and ROP37 (SRCN_7084), implying that the ROPKs are not restricted to *T. gondii*. Although largely presumed to be inactive, ROPKs are implicated in the regulation of the host transcription [48], and their presence in *S. neurona* may support the hypothesis that the ROPKs have a unique activation mechanisms in their regulatory functions that facilitate apicomplexan pathogenesis [35, 49]. Other notable OPKs included two parasite-specific eukaryotic initiation factor-2 (eIF2) kinases (eIF2K-C [SRCN_1606] and eIF2K-B [SRCN_4503]), four NEKs (SRCN_4528, SRCN_2630, SRCN_286 and SRCN_3151) and four ULKs (SRCN_3444, SRCN_3669, SRCN_6812 and SRCN_6157) (Table 1). The eIF2Ks are conserved in apicomplexans, and are important for the induction of parasite differentiation into the bradyzoites cysts, which are clinically important [29].

2.1.6 The STE group

The STEs are poorly represented in apicomplexans, and although most apicomplexans have one or two STE genes per genome, some parasites such as *C. parvum* are reported to harbor up to six STEs [15, 17]. Our results suggest that *S. neurona* has at least one putative STE (Table 1). STEs are thought to function in MAPK pathway cascades despite the fact that this pathway is absent in apicomplexans. The small repertoire of apicomplexan STEs is in contrast to that reported in other parasites such as trypanosomatids, in which these enzymes regulate the length of the flagella [50].

2.1.7 The TKL group

Apicomplexans harbor a maximum of seven TKL-coding genes, which makes it notable that we identified six putative TKLs in *S. neurona* (Table 1). In *Plasmodium*, six TKLs are conserved, and are essential in the asexual proliferation of the parasites, hence the TKLs are thought to be ideal drug targets [22]. Two of the six *S. neurona* putative TKLs had considerable sequence similarities to the *Plasmodium* TKLs, including SRCN_3466 (36% similar to *P. malariae* TKL1) and SCRN_1435 (49% similar to *P. ovale* TKL3) (data not shown).

2.1.8 The aPK group

The aPKs share sequence similarity to the OPKs. Although apicomplexan kinomes are generally thought to be devoid of aPKs, at least four genes in *T. gondii* are thought to encode these enzymes, the products of which are hypothesized to be part of the Ovoid Mitochondrial Cytoplasmic (OMC) Complex [51], a composite assembly of organelles observed only in growing tachyzoites of *T. gondii*, but its functions are currently unclear. Our analysis revealed two putative MEK-related aPKs in *S. neurona* (see Table 1). A representative of an MEK-related aPK (PfPK7) is reported to be expressed during *Plasmodium* infections in humans [52].

2.2 Evolution of *S. neurona* protein kinases

We phylogenetically deciphered evolutionary relationships among the various PK groups in the *S. neurona* kinomes with the PKs in other apicomplexans, which revealed that the *S. neurona* kinome has slightly fewer AGCs (*n*=9) compared to the kinomes of *T. gondii* (*n*=11), *N. caninum* (*n*=13) and *H. hammondi* (*n*=15). In general, the phylogenetic clustering of the *S. neurona* AGCs mirrored the homologies of these enzymes to those of the three apicomplexans used in this study (Figure 1; compare with Table 1).
Sequence analysis of *S. neurona* AGCs revealed significant divergence with only ~30% sequence similarity. Two *S. neurona* AGCs (SRCN_5610 and SRCN_3990) clearly cluster with *T. gondii* PKAs TGME49_028420 and TGME49_015670 [53] (Figure 1). Moreover, SRCN_5610 shares ~60% full length sequence identity with its ortholog. Further, two other *S. neurona* putative PKAs (SRCN_5165 and SRCN_3339), which also distinctly clustered with their *T. gondii* orthologs (TGME49_026030 and TGME49_086470, respectively) (Figure 1), contained the characteristic PKA’s ‘GxGxxG’ motif harboring a glycine triad that structurally forms a hairpin around the ATP binding pocket [53]. It is also notable that the single putative PKG (SRCN_4518) distinctly clustered with its *T. gondii* PKG ortholog (TGME49_111360) (Figure 1).

The SRCN_3339, SRCN_5165, SRCN_5610 and SRCN_4518 possessed an additional functional domain; the AGC-kinase C-terminal domain that contained two of the three conserved
phosphorylation sites in AGCs (data not shown). These conserved sites serve as phosphorylation-regulated switches to control both intra- and inter-molecular interactions [54]. A striking feature in *S. neurona* is that like in *T. gondii*, it lacks PKB and PKC. The putative PDK1 (SRCN_1312) clustered with the *T. gondii* PDPK (TGME49_068210) [53].

Despite the absence of PKC in *S. neurona*, CAMK family members were identified, which perhaps underscores the importance of Ca^{2+} regulation in this apicomplexan. Majority of the identified *S. neurona* CAMKs segregated with their orthologs in *T. gondii*, *N. caninum* and *H. hammondii* in clades with robust bootstraps (Figure 2), thus validating the annotation of the CAMKs. Amongst the CAMKs, SRCN_2544 clustered with *T. gondii* PK1 (TGME049_043500) of the AMPK/SNF1 sub-family. There were also three additional SNF1 members in *S. neurona* (SRCN_5410, SRCN_4815 and SRCN_2257), which clustered with *T. gondii* TGME49_115190, TGME49_033900 and TGME49_091050, respectively.

Figure 2. Mid-point rooted ML phylogenetic tree of apicomplexan CAMKs. The terminal branches are color-coded for AGCs in the kinomes of *S. neurona* (SRCN; red), *T. gondii*, ME-49 strain (TGME49; blue), *H. hammondii* (HHA; green) and *N. caninum*, Liverpool strain (NCLIV; purple). A solid purple circles on a branch indicates bootstrap support greater than 70. The phylogenetic tree was inferred from a multiple sequence alignment using PhyML with LG amino acid substitution model and gamma model of substitution rate heterogeneity. The tree image was rendered with iTOL.
Based on the clustering with *T. gondii* CDPK orthologs, 10 *S. neurona* CDPKs, including CDPK1 (SRCN_3314), CDPK2 (SRCN_4390), CDPK2B (SRCN_2165), CDPK3 (SRCN_3701), CDPK4 (SRCN_6606), CDPK5 (SRCN_3583), CDPK6 (SRCN_3011), CDPK7 (SRCN_6597), CDPK8 (SRCN_5948) and CDPK9 (SRCN_5812) were identified (Figure 2). This result implies that *S. neurona* has potentially lost at least two CDPKs (compared to the 12 CDPK that have been reported in *T. gondii* [21]). The possible loss notwithstanding, *S. neurona* contained the 6 CDPKs that are expressed and are well-conserved in most apicomplexans (i.e. CDPK1, CDPK3, CDPK4, CDPK5, CDPK6, and CDPK7) [55]. Sequence analysis revealed that, like in other apicomplexans, all identified *S. neurona* CDPKs except CDPK7 (SRCN_6597) contained both a kinase domain and a Ca\(^{2+}\)-binding domain known as the EF-hand domain [21]. Like its *T. gondii* ortholog, TGME49_028750 (TgCDPK7), the *S. neurona* CDPK7 (SRCN_6597) contains a pleckstrin-homology (PH) domain just upstream of its PK domain [55]. The domain architecture in CDPKs is such that kinase activity is stimulated upon Ca\(^{2+}\)-binding. Finally, the clustering of SRCN_5227, SRCN_1071, SRCN_3011 and SRCN_4093 in Figure 2 imply possible species-specific expansion of the CAMKs in *S. neurona*.

Majority of the putative CMGCs identified in *S. neurona* clustered with robust bootstraps with the conserved CMGCs in *T. gondii*, *N. caninum* and *H. hammondi* (Figure 3). Notable were the clustering of the CDKs, including CDK7 (SRCN_4674, SRCN_2759 and SRCN_761), CDK10 (SRCN_895) and CDK11 (SRCN_977). In addition, comparing the phylogenetic clustering of the CDKs in Figure 3 and the annotations presented in Table 1, the *S. neurona* kinome also appeared to have putative CDK5 (SRCN_4801, SRCN_1104 and SRCN_6346), all of which were supported by robust bootstraps. It should be noted that CDKs are amongst the principle molecular switches that regulate cell cycle progression in the apicomplexan parasites [56]. Also notable was the clustering in the same clade of three MAPKs (SRCN_6472, SRCN_4209 and SRCN_5365) and two GSKs (SRCN_1731 and SRCN_1732) (Figure 3).
Figure 3. Mid-point rooted ML phylogenetic tree of apicomplexan CMGCs. The terminal branches are color-coded for AGCs in the kinomes of S. neurona (SRCN; red), T. gondii, ME-49 strain (TGME49; blue), H. hammondi (HHA; green) and N. caninum, Liverpool strain (NCLIV; purple). A solid purple circle on a branch indicates bootstrap support greater than 70. The phylogenetic tree was inferred from a multiple sequence alignment using PhyML with LG amino acid substitution model and gamma model of substitution rate heterogeneity. The tree image was rendered with iTOL.

In the OPK family, SRCN_4528 and SRCN_2630 are putative NEKs given their clustering with T. gondii NEK kinases TGME49_119700 and TGME49_094260, respectively (Figure 4). The S. neurona SRCN_108 and SRCN_3444 are putative ULKs due to their clustering with their T. gondii ULK orthologs (TGME49_035750 and TGME49_040630, respectively). SRCN_3669 clustered with TGME49_066950, a TBC-domain (Tre-2/Bub2/Cdc16) containing kinase. Further, there seemed to be a species-specific expansion of S. neurona Aurora kinases, SRCN_1606, SRCN_6157 and SRCN_2403 which cluster with T. gondii aurora kinase TGME49_003010; the clade containing these kinases however was not supported by robust bootstrap values (see Figure 4). SRCN_286 is a putative Wee kinase given that it clusters with Wee kinases from T. gondii (TGME49_073690) as well as from H. hammondi (HHA_273690; compare Figure 4 with Table 1).
Figure 4. Mid-point rooted ML phylogenetic tree of apicomplexan OPKs. The terminal branches are color-coded for AGCs in the kinomes of *S. neurona* (SRCN; red), *T. gondii*, ME-49 strain (TGME49; blue), *H. hammondi* (HHA; green) and *N. caninum*, Liverpool strain (NCLIV; purple). A solid purple circle on a branch indicates bootstrap support greater than 70. The phylogenetic tree was inferred from a multiple sequence alignment using PhyML with LG amino acid substitution model and gamma model of substitution rate heterogeneity. The tree image was rendered with iTOL.

The clustering of putative *S. neurona* ROPKs (Figure 4) was also notable. For instance, SRCN_7082 clustered with *T. gondii* ROP33 (TGME49_001130), implying that it is a ROP33 (Table 1). SRCN_6184 clustered with *T. gondii* ROP19A (TGME49_042240) and ROP38 (TGME49_042110). It should however be noted that the sequence similarities with the ROP38 was low, implying that this could be a ROP19A (based on annotations presented in Table 1). Other putative *S. neurona* ROPKs included SRCN_3247, which clustered with ROP27 (TGME49_113330), SRCN_2076 with ROP25 (TGME49_002780). Here, it is notable that the clustering of SRCN_2076 the *T. gondii* ROP25 was supported with low bootstrap value, potentially meaning that based on the annotations (Table 1), SRCN_2076 is an ortholog to ROP30. SRCN_2123 and SRCN_2183 clustered with ROP35 (TGME49_104740), while SRCN_7084 clustered with ROP37 (TGME49_094560). Notably, from the phylogenetic tree in figure 4, SRCN_7083 and SRCN_4410 are probably duplicated forms of ROP34 given their clustering with *T. gondii* TGME49_040090. There was also indication of species-expansion of the *S. neurona* putative ROP35 (SRCN_2123 and SRCN_2183; see Figure 4). Although annotations
(Table 1) showed that SRCN_4310 was a putative ROP33 (39% sequence similarity to *H. hammondi* ROP33), this enzyme clustered with a hypothetical protein (TGME49_039380) of *T. gondii* (Figure 4). This discrepancy between annotation and phylogeny could be due to differences in the *T. gondii*. Another apparent discrepancy was the annotation of SRCN_4503 as a putative elf2K-B (74% sequence similarity to the *T. gondii* TgCatPRC2 [IF2K-B]; see Table 1); but phylogeny showed clustering of this protein with a putative *T. gondii* ROP30 (TGME49_027010) (see Figure 4). Note that in this figure, the phylogeny bootstraps were low in the segregation of SRCN_4503 with ROP30. Moreover, SRCN_3151 and SRCN_3075 both segregated with TGME49_018550, a PIK3R4 kinase-related protein; from the phylogeny, SRCN_3151 and SRCN_3075 probably represent duplications. SRCN_6572 falls in the same clade with TGME49_004100, an elf2 kinase IF2K-C, and TGME49_111510, an elf2 kinase IF2K-B. NEKs are involved in cell cycle regulation, while Aurora kinases play pivotal roles in endodyogeny, duplication rate and parasite virulence [28]. Taken together, the presence of a variety of ROPKs in *S. neurona* is interesting given the fact that in *T. gondii*, ROPKs are key virulence factors [57].

### 2.3 Structural modeling

Using I-TASSER, we constructed the three-dimensional (3-D) homology models for SRCN_3247 using 4B6L, 4C0T, 1KOB, 4ZHX, 4IC7 and 4CZU as templates, and SRCN_5165 using 1CDK, 1SMK, 4WH, 4Z84, 3PFQ, 4LQS and 1RDQ as templates. SRCN_3247 and SRCN_5165 are putative ROP27 and PKA respectively. From the predicted models for each of the sequences, the model with the highest C-score, a score that estimates the quality of the models produced, was used. The best SRCN_3247 model had a C-score of -0.46 and an estimated TM-score of 0.65 (Figure 5A) while the best SRCN_5165 model had a C-score of 1.6 and a TM-score of 0.94 (Figure 5B).

![Figure 5. (A) Homology model of S. neurona SRCN_3247 (ROP27) generated based on template structures 4B6L, 4C0T, 1KOB, 4ZHX, 4IC7 and 4CZU. (B) Homology model of SRCN_5165 generated based on template structures 1CDK, 1SMK, 4WH, 4Z84, 3PFQ, 4LQS and 1RDQ.](https://example.com/fig5.png)

Evaluation of the models’ quality and suitability for docking using PROCHECK revealed that SRCN_3247 and SRCN_5165 had 97.4% and 99.7% of the residues in the favored/allowed regions, respectively (Supplementary Figure 1A&B). The SRCN_3247 model contained 9 beta-sheets, 11 alpha and 5 3,10 helices while SRCN_5165 contains 9 beta-sheets, 11 alpha and 6 3,10 helices. Using the improved Protein Block Alignment (iPBA) webserver [58], structural superposition of the models to the top-ranking templates was performed for prediction of pair-wise structural alignment. The structural similarity score was measured using root mean square deviation (RMSD) between both Cα atom of the modeled protein and the corresponding template. The RMSD for the alignment of SRCN_3247 and 4B6L was 0.82 Å for the 276 aligned residues (Figure 6A) while that for
SRCN_5165 and 1CDK was 0.46 for 335 aligned residues (Figure 6B). The obtained RMSD values indicated that in both cases, the models and the templates had similar folds.

![Figure 6A](image1.png) ![Figure 6B](image2.png)

**Figure 6.** A. Superimposition of SRCN_3247 (marine) and 4B6L template (red). B. Superimposition of SRCN_5165 (purple) and its template, 1CDK (red).

Analysis of putative binding pockets using DoGSiteScorer [59] revealed 14 potential pockets (P0-P13) on the surface of the SRCN_3247 model and 13 potential pockets (P0-P12) on the surface of the SRCN_5165 model. The largest pocket in SRCN_3247 (P0; Figure 7A) had a volume of 624.06 Å³ and a drug score of 0.79 while the largest pocket in SRCN_5165 (P0; Figure 8) had a volume of 760.51 Å³ and a drug score of 0.77.

![Figure 7A](image3.png) ![Figure 7B](image4.png)

**Figure 7.** A cartoon of the largest binding pockets of SRCN_3247 (Panel A) and SRCN_5165 (panel B).

Here, it should be noted that the presence of a binding pocket does not necessarily imply that a target protein is druggable. However, the binding volume is one of the important cavity properties
that influence the druggability of a particular target protein, of which most of druggable proteins have volumes of between 500-1000 Å³ [60].

2.4 Molecular docking of SRCN_3247 and SRCN_5165 with PK inhibitors

After the structural modeling, the binding modes of known kinase inhibitors to SRCN_3247 and SRCN_5165 was assessed by molecular docking of the chemical ligands into the predicted active site pockets of the modeled structures. For the docking analyses, the following four inhibitors were used: - BAY61-3606, an ATP-competitive inhibitor of mammalian spleen tyrosine kinase (Srk) [61]; Flavopiridol, a synthetic flavonoid that inhibits a wide range of cyclin-dependent kinases [62]; purvalanolB-B (PurB) is a potent CDK inhibitor [63]; and Piceatannol, a naturally occurring hydroxylated analogue of resveratrol that inhibits Syk kinase [64]. With a binding affinity of -8.4 kcal/mol, the overall binding pose of BAY61-3606 in SRCN_5165 is shown in Figure 8A. The binding affinity of flavopiridol to SRCN_5165 was -9.3 kcal/mol (Figure 8B). The binding energies for purB and piceatannol to SRCN_5165 were -6.1 kcal/mol (Figure 8C), and -6.0 kcal/mol (Figure 8D), respectively.

Figure 8. (A) Binding of BAY61-3606 to the active site pocket of SRCN_5165; (B) Binding of flavopiridol to the active site pocket of SRCN_5165; (C) Binding of purvalanolB to the active site pocket of SRCN_5165; and (D) Binding of piceatannol to the active site pocket of SRCN_5165. The inhibitors and structural models are shown in stick and cartoon representations, respectively.
For SRCN_3247, BAY-61-3606 bound to the active site with an affinity of -7.2 kcal/mol (Figure 9A), while the affinities for flavopiridol, purB and piceatannol were -7.9 kcal/mol (Figure 9B), -6.9 kcal/mol (Figure 9C), and -7.3 kcal/mol (Figure 9D), respectively.

![Figure 9](image)

Figure 9. (A) Binding of BAY61-3606 to the active site pocket of SRCN_3247; (B) Binding of flavopiridol to the active site pocket of SRCN_3247; (C) Binding of purvalanolB to the active site pocket of SRCN_3247; and (D) Binding of piceatannol to the active site pocket of SRCN_3247. The inhibitors and structural models are shown in stick and cartoon representations, respectively.

Taken together, by structural modeling and docking analysis, the above results indicate that the putative *S. neurona* PKA and ROP27 could be considered as drug targets. The data from the modeling and docking could pave way for further laboratory experiments to design potential drugs against EMP.

3. Discussion

We used different bioinformatics approaches to identify and classify the PKs of *S. neurona*, and docked four well-known inhibitors into modelled structures of two PK representatives. The critical roles performed by PKs make them attractive targets for novel therapeutics, and as such, the current study provides a rich source of potential drug targets, vaccine candidates or diagnostic proteins for developing new treatment and diagnostics strategies. PKs are considered viable drug targets because their catalysis mechanism and overall structure are conserved. Because small molecules can bind to the PKs’ catalytic clefts [65], many kinase inhibitors (e.g. imatinib, trastuzumab and lapatinib) have been developed to treat various human diseases [66].
The kinomes of apicomplexans range from 35 PKs (in Babesia bovis) to 135 PKs (in T. gondii) [35]. We identified a total of 92 putative PKs in the kinome of S. neurona, compared to the PKs reported in the kinomes of T. gondii (n=135), N. caninum (n=130) and H. hammondi (n=124) [15]. Although the total number of S. neurona PKs appeared markedly reduced compared to that of its close coccidian relatives (T. gondii and N. caninum [27]), taken as a percentage of total genome size, the proportion of S. neurona PKs is comparable to the 2% observed in humans [13] and other coccidians [27]. The contraction of the S. neurona kinome could be attributed to genome compaction, which occasionally offsets lineage-specific expansions of specific gene families. Notably, genome contraction is a common mode of genomic evolution in intracellular parasites, including apicomplexans [67, 68]. As such, the evolution of PKs may be in tandem to the overall genomic adaptive strategies of these parasites.

Using a hierarchical scheme based on the major PK groups, the S. neurona kinases could be classified and phylogenetically clustered into the various PK families. A complement of nine putative AGC kinases was identified in S. neurona, which is reduced compared with that of T. gondii, N. caninum and H. hammondi. Despite this potential gene loss, seven of the nine AGCs (SRCN_5165, SRCN_3339, SRCN_4249, SRCN_3990, SRCN_5430 and SRCN_1312) had orthologs in T. gondii, N. caninum and H. hammondi. In agreement with the observation that PKA is conserved in apicomplexans [27], two PKAs (SRCN_5610 and SRCN_3990) were identified in S. neurona. In eukaryotes, PKAs signaling regulate various cellular responses such as DNA replication [69], cell growth and metabolism [70] and gene transcription by phosphorylating transcription factors such as the cAMP Response Element-Binding Protein (CREB) [71]. In T. gondii, increases in cytosolic cAMP levels activate PKA to trigger the developmental switch from the rapidly proliferating tachyzoites to the quiescent bradyzoites [72]. Additionally, two other S. neurona AGCs, (SRCN_5165 and SRCN_3339), were putative PKAs given that they contained the characteristic GxGxxG motif found in PKA [53]. The PKAs are attractive targets for the disruption of S. neurona growth. Notably, based on orthology, S. neurona contains a single putative PKG (SRCN_4518) that distinctly clustered with T. gondii PKG (TGME49_111360).

In apicomplexans, CAMKs modulate the intracellular Ca²⁺ concentration, which in turn regulates vital processes such as host-cell invasion, protein secretion and parasite differentiation. We identified four AMPK/SNF1 family members (SRCN_2544, SRCN_5410, SRCN_4815 and SRCN_2257). The AMP-activated PK cascade acts as a metabolic sensor that monitors cellular AMP and ATP levels and is activated by an elevation of the AMP:ATP ratio. In yeast, scarcity of nutrients switching off ATP-consuming anabolic pathways and switching on ATP generating catabolic pathways such as fatty acid oxidation [74]. Further, we identified 10 putative CDPKs in the S. neurona, including CDPK1 (SRCN_3314), CDPK2 (SRCN_4390), CDPK2A (SRCN_2165), CDPK3 (SRCN_3701), CDPK4 (SRCN_6606), CDPK5 (SRCN_3583), CDPK6 (SRCN_3011), CDPK7 (SRCN_6597), CDPK8 (SRCN_5948) and CDPK9 (SRCN_5812). Compared to the 12 CDPKs reported in T. gondii [21], it appears that S. neurona had all the six well-conserved apicomplexan CDPKs (CDPK1, CDPK3, CDPK4, CDPK5, CDPK6 and CDPK7), which provide a link between Ca²⁺ signaling and parasite differentiation, motility, invasion and egress [55]. In T. gondii, down-regulation of CDPK1 interfered with parasite motility, host cell invasion and egress [44] while disruption of CDPK3 caused defective parasite egress [75]. Further, the essentiality of CDPK6 and CDPK7 in T. gondii have recently been demonstrated [76]. Indeed, TgCDPK1 has been targeted for the development of new drugs for toxoplasmosis [77]. Sequence analysis revealed that, similar to other apicomplexans, all identified S. neurona CDPKs except CDPK7 (SRCN_6597) contain both a PK domain and a EF-hand (Ca²⁺-binding) domain [21]. Similar to its T. gondii ortholog, TGME49_028750 (TgCDPK7), the S. neurona CDPK7 (SRCN_6597) contains a pleckstrin-homology (PH) domain just upstream of its PK domain [55]; the domain architecture is such that kinase activity is stimulated upon Ca²⁺ binding. Moreover, our phylogeny provided clues of possible gene duplications giving rise to SRCN_3990 and SRCN_3011 and well as SRCN_4093 and SRCN_1071). Interestingly, based on phylogenetic analysis, S. neurona probably contains four (SRCN_5227, SRCN_1071, SRCN_4093 and SRCN_3011) species-specific CAMKs.
The CMGCs, comprising of CDKs, MAPKs, GSKs, and CLKs coordinate a wide range of cellular functions in different species. For instance, members of the CDK subfamily are major coordinators of cell division in both mitosis and meiosis while the MAPKs subfamily is crucial in regulating cell proliferation, cell differentiation and cell death in eukaryotes [56, 78]. By both annotations and phylogenetic analyses, we identified four putative CDKs sub-family members; CDK5 (SRCN_4801 and SRCN_6346), CDK7 (SRCN_2759, SRCN_4674 and SRCN_761), CDK10 (SRCN_895) and CDK11 (SRCN_977). The finding of CDKs in *S. neurona* suggests that this parasite’s cell cycle regulation could be CDK-dependent and perhaps similar to that of higher eukaryotes [79]. The identification of three putative MAPKs (SRCN_4209, SRCN_6472 and SRCN_5365) in *S. neurona* points to the existence of MAPK regulated transduction pathway(s) in this pathogen. Similar to its *T. gondii* ortholog (TgMAPK1), which is a p38α MAPK homolog [80], SRCN_4209 may be potentially involved in parasite proliferation/stage differentiation, stress response and manipulation of the host immunity to enhance virulence. On the other hand, SRCN_6472 and SRCN_5365 may augment the roles of SRCN_4209 in the parasite. In *T. gondii* MAPK1/ERK7 is involved in intracellular proliferation [81]. We also identified two putative GSKs (SRCN_1731 and SRCN_1732). A genome-wide gene knockout approach in *P. falciparum* demonstrated PfGSK-3 to be critical for schizogony of the parasite [82]. Other *S. neurona* CMGC kinases identified include CLK (SRCN_1479), PRP4 (SRCN_2845), DYRK (SRCN_1611) CK2 (SRCN_6427) and SRPK (SRCN_1236).

ROPKs are secreted by *T. gondii* into the host cell and play roles in adhesion, motility and manipulation of immune responses [83]. We identified at 10 putative ROPK sub-family members in the *S. neurona* kinome; - ROP38 (SRCN_6184), ROP33 (SRCN_7082), ROP27 (SRCN_3247), ROP25 (SRCN_2076), ROP37 (SRCN_7084), ROP34 (SRCN_4410 and SRCN_7083), ROP35 (SRCN_2183), SRCN_2123 and ROP23 (SRCN_6812). It has been recently demonstrated that *T. gondii* ROP21 and ROP27 play a role in a constitutive pathway based on their localization in the PV and cyst matrix [84]. Moreover, *T. gondii* ROP35 has been shown to play a crucial role in chronic infection [85]. Although the *S. neurona* genome is more than twice the size of other coccidians whose genomes sequenced so far (e.g. *Toxoplasma* and *Neospora*), it has a considerably reduced number of ROPKs that nevertheless may have vital roles in the parasite’s virulence. Specifically, *S. neurona* is devoid of ROP5, ROP16, ROP18 and ROP38 that have been shown to confer virulence and alter the host’s cellular signaling pathways [81]. Putatively therefore, *S. neurona* ROPKs may have multiple roles in the survival of the parasite. In search of drug targets against *S. neurona*, the reduced ROPKs with possible multiple roles and absent in the vertebrate host are thus attractive candidates.

We explored the possibility of repositioning existing PK inhibitors as lead compounds in the development of new drugs for EPM, an approach that would substantially lower the cost of drug development. The lead compounds found via repositioning could be modified to increase efficacy. Molecular docking studies were therefore conducted to explore the potential binding of four inhibitors to the structural models of SRCN_3247 and SRCN_3314, representatives of putative *S. neurona* ROP27 and PKA, respectively. Probing the *S. neurona* kinome with small-molecule has many potential advantages. For instance, given that most kinase inhibitors in therapeutic use target multiple kinases with similar binding pockets [65], the promiscuity of such agents lead to polypharmacology [86] that would help minimize drug resistance. However, our modeling and docking analysis are only predictive, meaning that experimental validations are required.

4. Conclusion

The kinome of *S. neurona* contain members of the major classes of PKs, including AGC, CMGC, GSK, CAMK, CK, TKL, aPKs and several PKs in the OPK family. Similar to other apicomplexans, *S. neurona* kinome is devoid of PKC and the conventional TKs, but appears to have reduced MAPK family members. *S. neurona* kinome did not also have some of the ROPKs that have been implicated in the virulence of *T. gondii*. Known kinase inhibitors can be used as starting scaffolds in the search for drugs acting against EPM.
5. Materials and Methods

5.1. Genome-wide identification of putative *S. neurona* PKs

The predicted *S. neurona* proteome was downloaded from Toxoplasma Genomics Resource database (Release 28; version May 2016) [41]. A hidden Markov model (HMM) profile of signature PK domains obtained from the Kinomer database v 1.0 [87] was used to search for *S. neurona* kinases using HMMER v 3.1b2 [88]. The sequences having PK domain (IPR011009) or PK-like domain (IPR000719) were considered as putative kinases. Annotation of the putative kinase sequences was performed by blast search against the non-redundant (nr)-NCBI protein and UNIPROT/SWISSPROT databases at an e-value of 0.0001. Gene ontology (GO) mapping was performed using Blast2GO v 3.3 [89]. The molecular weight (Mw) and isoelectric point (pI) were obtained using ExPASy compute pI/Mw tool [90]. Motifs analysis was performed with the MEME Suite v 4.11.2 [91]. The parameters were as follows: number of repetitions, any; maximum numbers of motifs, 30; and the optimum motif widths, between 6 and 200 residues.

5.2. Phylogenetic analysis

Phylogenetic trees were constructed to decipher the orthologous and paralogous relationships of *S. neurona* kinases. Sequences of *S. neurona* kinase subfamilies obtained using Blast2Go were aligned with their homologs in *T. gondii* [41], *H. hammondi* and *N. caninum* using MUSCLE [92]. The alignments were subsequently manually edited in Jalview [93]. Phylogenetic reconstruction was undertaken using the maximum likelihood program PhyML 3.0 [94] and RAxML v 8.0 [95] and the Bayesian inference program MrBAYES v 3.2 [96]. For PhyML, the LG substitution model was selected assuming an estimated proportion of invariant sites and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data. Robustness of internal branches was evaluated using 100 bootstraps. MrBayes was run for 5,000,000 generations with two runs and four chains in parallel and a burn-in of 25%. Obtained trees were rendered with the Interactive Tree of Life server (iTOL) [97].

5.3. Structural modeling

The amino acid sequences of the target kinases, a putative ROP27 (SRCN_3247) and a PKA (SRCN_5165) (see Table 1), were retrieved ToxoDB [30]. Structural modelling was carried out using two approaches. The first approach was modelling by satisfaction of spatial restraints as implemented in MODELLER [98]. For this approach, template structures were obtained by threading target sequences in the PHYRE database [99] and selecting top ranked template structures for model building. The second modelling approach was the iterative threading assembly refinement implemented in I-TASSER [100]. The template structures for SRCN_3247 were 4B6L, 4COT, 1KOB, 4ZHX, 4IC7 and 4CZU, while those for SRCN_5165 were 1CDK, 1SMK, 4WH, 4Z84, 3PFQ, 4LQS and 1RDQ. The quality of the models was further evaluated using PROCHECK [101] to ensure suitability for carrying out docking studies. PROCHECK outputs a Ramachandran plot that shows the number (%) of residues mapping to the favored, allowed, generally allowed and disallowed regions of the plot. Structural superpositioning was performed using the improved Protein Block Alignment (iPBA) webserver [58] Results were visualized in the PyMOL Molecular Graphics System [102].

5.4 Molecular docking

Docking calculations were carried out using the AutoDock Suite [103]. To compute the free energy of binding on each protein model, essential hydrogen atoms and solvation parameters were added. Dockings were performed with grid boxes sufficient to cover the entire protein of interest using parameters tabulated in Table 2. Four PK inhibitors, BAY61-3606, Flavopiridol, Purvalanol B
and Piceatannol were docked onto SRCN_3247 and SRCN_5165 and their overall binding poses evaluated in terms of binding energies.

Table 2. Parameters used for molecular docking of kinase inhibitors with SRCN_3247 and SRCN_5165, which are representatives of S. neurona putative ROP27 and PKA, respectively.

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Ligands</th>
<th>Center-X</th>
<th>Center-Y</th>
<th>Center-Z</th>
<th>Size-X</th>
<th>Size-Y</th>
<th>Size-Z</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRCN_5165</td>
<td>BAY61-3606</td>
<td>65</td>
<td>63</td>
<td>60</td>
<td>52</td>
<td>62</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Flavopiridol</td>
<td>63.929</td>
<td>62.41</td>
<td>65.345</td>
<td>56</td>
<td>58</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Purvalanol B</td>
<td>62.0</td>
<td>60.5</td>
<td>63.0</td>
<td>50</td>
<td>55</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>Piceatannol</td>
<td>58</td>
<td>61.23</td>
<td>63</td>
<td>48</td>
<td>50</td>
<td>58</td>
</tr>
<tr>
<td>SRCN_3247</td>
<td>BAY61-3606</td>
<td>62.665</td>
<td>64.359</td>
<td>65.712</td>
<td>62</td>
<td>48</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Flavopiridol</td>
<td>60.157</td>
<td>64.359</td>
<td>65.52</td>
<td>60</td>
<td>55</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Purvalanol B</td>
<td>62.34</td>
<td>60.15</td>
<td>63.45</td>
<td>56</td>
<td>55</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Piceatannol</td>
<td>63.23</td>
<td>62.43</td>
<td>60.15</td>
<td>60</td>
<td>56</td>
<td>45</td>
</tr>
</tbody>
</table>

All grid boxes were formed at spacing size of 1.0 Angstrom. Following the docking procedure, the resulting PDBQT output files were visualized using PyMOL [102] and all ligand conformations were analyzed for their binding affinities.

Supplementary Materials

Supplementary Figure S1: Ramachandran plots generated using PROCHECK to evaluate the robustness three-dimensional models SRCN_5165 (Panel A) and SRCN_3247 (Panel B), which are putative PKA and ROP27, respectively, in S. neurona kinome. Here, the red, yellow, light yellow, white regions indicate the favored, allowed, generously allowed and disallowed regions, respectively. The Phi and Psi are dihedral angles that determine torsion of peptide bonds.

Acknowledgments

The authors thank Dr. Peter Waweru for useful discussions and Egerton University for funds to cover the costs to publish in open access.

Author Contributions

E.M conceived and designed the experiments; E.M and H.K performed the experiments, analyzed the data and wrote the paper. The authors read an approved the final version of the manuscript.

Conflicts of Interest

The authors declare that there is no conflict of interest in this work.

References


© 2017 by the authors; licensee *Preprints*, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons by Attribution (CC-BY) license (http://creativecommons.org/licenses/by/4.0/).