

# A Novel Predicted Calcium-Regulated Kinase Family Implicated in Neurological Disorders

Małgorzata Dudkiewicz<sup>1</sup>, Anna Lenart<sup>2</sup>, Krzysztof Pawłowski<sup>1,2\*</sup>

<sup>1</sup> Warsaw University of Life Sciences, Warsaw, Poland, <sup>2</sup> Nencki Institute of Experimental Biology, Polish Academy of Sciences, Warsaw, Poland

## Abstract

The catalogues of protein kinases, the essential effectors of cellular signaling, have been charted in Metazoan genomes for a decade now. Yet, surprisingly, using bioinformatics tools, we predicted protein kinase structure for proteins coded by five related human genes and their Metazoan homologues, the FAM69 family. Analysis of three-dimensional structure models and conservation of the classic catalytic motifs of protein kinases present in four out of five human FAM69 proteins suggests they might have retained catalytic phosphotransferase activity. An EF-hand Ca<sup>2+</sup>-binding domain in FAM69A and FAM69B proteins, inserted within the structure of the kinase domain, suggests they may function as Ca<sup>2+</sup>-dependent kinases. The FAM69 genes, *FAM69A*, *FAM69B*, *FAM69C*, *C3ORF58 (DIA1)* and *CXORF36 (DIA1R)*, are by large uncharacterised molecularly, yet linked to several neurological disorders in genetics studies. The *C3ORF58* gene is found deleted in autism, and resides in the Golgi. Unusually high cysteine content and presence of signal peptides in some of the family members suggest that FAM69 proteins may be involved in phosphorylation of proteins in the secretory pathway and/or of extracellular proteins.

**Citation:** Dudkiewicz M, Lenart A, Pawłowski K (2013) A Novel Predicted Calcium-Regulated Kinase Family Implicated in Neurological Disorders. PLoS ONE 8(6): e66427. doi:10.1371/journal.pone.0066427

**Editor:** Jonathan Wesley Arthur, Children's Medical Research Institute, Australia

**Received:** December 20, 2012; **Accepted:** May 8, 2013; **Published:** June 28, 2013

**Copyright:** © 2013 Dudkiewicz et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** These authors have no support or funding to report.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: krzysztof\_pawlowski@sggw.pl

## Introduction

In pre-genomics biology, defined molecular function in search of a protein effector was a common reality. The opposite, a protein in search of a function is still an even more common problem in biology.

Protein kinase-like (PKL) proteins are a huge grouping of regulatory/signalling and biosynthetic enzymes [1,2], regulating most processes in a living cell, by phosphorylating various substrates. Besides PKL kinases, other kinase families are known, functionally related, but of dissimilar structures [3]. Most PKL proteins feature a well-conserved structural scaffold, and a conserved active site [4,5]. These classic protein kinases number more than 500 in the human genome [1], and are among the most popular drug targets [6]. Within the protein kinase-like clan, sequence similarities between some families are relatively low. There are also PKL-like families that may be assigned to the clan mostly by the virtue of structural similarity. Despite the high interest in kinases (and overall almost six hundred thousand PubMed articles as of September 2012), the research effort has been biased [7,8] whereas approx. 10% of known kinases yielded at least 90% of publications [9]. Also, the human, and more generally, Metazoan kinome may be not fully charted yet. For example, a protein kinase-like domain has been recently discovered in selenoprotein O [10]. Also, a novel protein kinase family, FAM20, has been identified and characterised, [11–14], involved in phosphorylation of secreted proteins.

High-throughput studies often lead to discovery of disease links for uncharacterised proteins that due to lack of molecular function hypotheses are not followed upon. Here, using bioinformatics approaches, we analyse an obscure group of human proteins with

disease implications. First, we prove that FAM69 family members are homologous to protein kinases. Second, we predict that most FAM69s do have protein kinase activity. Third, we predict that some FAM69s are directly regulated by calcium ions via an EF-hand domain inserted in the middle of the kinase domain and close to the ATP-binding site. Fourth, we hypothesize that FAM69s may in fact regulate secretory pathways.

## Results

### FAM69s Belong to the Protein Kinase-like Clan

Examination of protein families with sequences distantly similar to the classical protein kinases, according to results of the FFAS algorithm [15], suggested that the uncharacterised Pfam family PIP49\_C (Pfam:PF12260, Pancreatitis induced protein 49 C terminal) [16] might be homologous to protein kinases. Indeed, using a member of the PIP49\_C family, the human FAM69A protein as a query, in the second PSI-Blast iteration [17] one detects significant sequence similarity to a human Ser/Thr protein kinase PKDCC (gi:292495024), with a significant E-value parameter equal to 3E-07, although with only 21% sequence identity over 197 residues. The distant albeit significant sequence similarity to both Ser/Thr kinases and to Tyr kinases does not allow for an unequivocal distinction between these two possibilities. Also, FFAS and HHpred structure prediction algorithms provide highly significant kinase-like predictions: Zscore -20 and E-value 1E-26, respectively. Among the structural predictions, Ser/Thr kinases are most common. The significant sequence similarity covers approximately the region 225–415 of the human FAM69A protein, that aligns to the C-terminal lobe of the kinase domain and the  $\beta$ -4  $\beta$ -5 region of the N-terminal lobe.

Since no equivalent of the ATP-binding glycine-rich loop which is typical for kinases was found in FAM69s, the N-terminal region of FAM69 was analysed separately. The region between residues 75 and 130 in human FAM69A exhibited sequence similarity of borderline significance to N-terminal lobe of protein kinases (FFAS Zscore  $-7.6$  and HHpred P-value  $1E-05$ ). Secondary prediction for this region supported the 3D structure prediction.

CLANS clustering analysis [18], presenting in a graph form distant sequence similarity relationships between groups of proteins, reflects relatively close similarity of FAM69s to many known kinase families (Fig. 1). Specifically, regardless of the sequence similarity clustering threshold used, FAM69s remain grouped with the central group of kinase families, including the classic pkinase and pkinase\_Tyr (Pfam database identifiers: PF00069, PF07714) families. Of note, some of the kinase families, members of the PKL clan for which homology to typical PKL proteins has been established, e.g. UL97 (PF06734), are found to be more distant from the central families than FAM69s (see Fig. 1).

The FAM69 family is present in most Metazoan taxa, with the sea anemone *Nematostella vectensis* being the organism most distant from humans possessing it. *N. vectensis* has four FAM69 genes, three of which code proteins with the typical FAM69 features conserved (See Fig. 2, left). FAM69s can be clearly separated into two major branches, one containing human FAM69A, FAM69B, FAM69C and the other containing C3ORF58 (DIA1) and CXORF36 (DIA1R) (See Fig. S1). The two protein groups have been analysed bioinformatically [19,20], but they have never been identified as homologues before, although their sequence similarity to each other is very significant (FFAS Zscore equals  $-56$  and HHalign E-value equals  $3E-12$  between human FAM69A and C3ORF58, see also Fig. 2, left).

### Most FAM69s are Predicted to be Active Kinases

Structure predictions are not automatically extendable to function predictions. Here, we will look more closely at the sequence and structure motifs important for kinase function.

FAM69s exhibit strong and significant sequence similarity to the C-terminal lobe of the classical kinase structure. Sequence similarity of FAM69s to the N-terminal smaller lobe of kinases is of borderline significance. Yet, after removal of the insertion sequence forming the EF-hand motif (approx. residues 165–190, see the following section), the remaining regions of FAM69A exhibit significant overall similarity to protein kinases (HHpred E-values below  $1E-33$ ).

Several of the conserved kinase regions, as defined for the archetypical protein kinase A (PKA) [21], are clearly conserved in FAM69s (See Fig. 2, left). The presumed catalytic base (D166 in PKA) is present in most FAM69s, excluding CXORF36, albeit in a [LM][CL]D motif (D294 in human FAM69A) instead of [HY]RD. Also well-conserved are residues corresponding to N171 and D184 of PKA, responsible for binding the  $Mg^{2+}$  ions (N299 and D312 in human FAM69A). Then, conserved are residues corresponding to K72 and E91 in strand  $\beta$ -3 and helix  $\alpha$ -C, respectively (K113 and E151 in human FAM69A) involved in binding the phosphate groups of the ATP molecule (see sequence logos in Fig. 3). However, the identification of the location of helix  $\alpha$ -C is less reliable than the identification of  $\beta$  strands 1, 2 and 3. As another possibility, the helical region following the predicted helix  $\alpha$ -C (see next section) may in fact provide an alternative location for a glutamate residue corresponding to the E91 of PKA kinase.

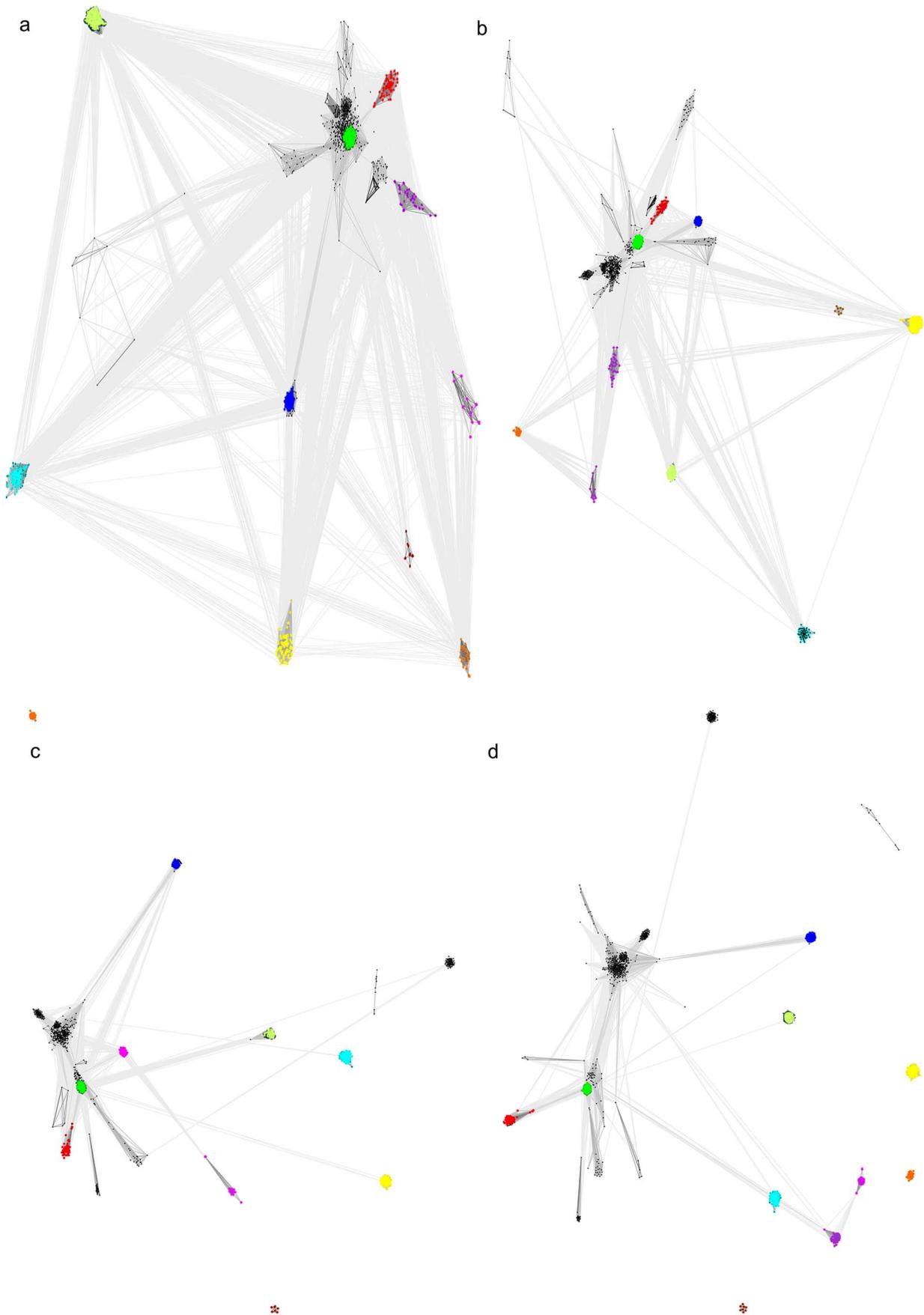
In cases of remote sequence similarity, building a structural model is of an illustrative nature rather than predictive, yet it serves also as a feasibility check for the predicted structure. We

analysed a structural model of the FAM69 kinase domain (residues 77–423, including the EF-hand domain region) built using selected kinase structures as templates (see Methods) and a manually curated sequence alignment (See Fig. S2). Overall, the structure model is reasonable as judged by the MetaMQAP model quality scoring (GDT\_TS parameter equal to 42.7 and expected C $\alpha$  RMSD from native structure is 4.8 Å).

The core structure of the protein kinases contains an unique ATP-binding motif, GxGxxG. In FAM69s, structure prediction algorithms do not allow unambiguous detection of a corresponding site. However, an analysis of weak fold predictions (FFAS and HHpred, see Methods), together with secondary structure predictions allowed detection of similarity to the kinase N-terminal lobe. Since the typical kinase Gly-rich motif is not present in FAM69s in the predicted  $\beta$ -1– $\beta$ -2 loop, it is likely that an atypical ATP-binding mode is employed in the FAM69 family, possibly using the partly conserved proline residue present in that loop in FAM69A and FAM69B, or the glycine residue that replaces it in FAM69C homologues.

The recently identified kinase features, the regulatory and catalytic spines built of conserved hydrophobic residues [22], are not easily identified in full in FAM69s due to substantial sequence divergence. Yet, most of these residues can be tentatively identified. In the regulatory spine, only L95 (PKA numbering) of helix  $\alpha$ -C cannot be easily identified in FAM69s. Otherwise, regulatory spine residues are conserved (see Fig. 2, left), including L106 of strand  $\beta$ -4, Y164 next to the predicted active site Asp, and F185 next to the  $Mg^{2+}$ -binding Asp184. Among the catalytic spine residues, V57 of strand  $\beta$ -2 is conserved, as well as A70 of strand  $\beta$ -3 and L172 and I174 located next to the predicted active site. The neighbouring L173 is usually not conserved as a hydrophobic residue. Finally, the M228 and M231 residues of helix  $\alpha$ -F are a difficult case due to the relatively uncertain sequence alignment for this region. However, interestingly, provided our alignment is realistic, these two residues are replaced by cysteines in FAM69A, B and C homologues which opens up a possibility that a disulphide bridge between these two residues stabilises the catalytic spine (See Fig. 2, left). Structural predictions for uncharacterised proteins are an important approach for advancing molecular biology and suggesting specific experimental approaches guided by the predictions [23–25]. Such approaches have been repeatedly successful. For example, the CLCA family was predicted to possess peptidase function, and later confirmed experimentally [26,27]. Likewise, the NLRP proteins were predicted to be involved in protein-protein interactions in immune responses and apoptosis, which has been subsequently validated [28,29]. The usefulness of a structural prediction depends on both the accuracy and the purpose of the prediction. It has been shown that inclusion of explicit water molecules greatly improves the quality of structural models and information derived therefrom [30,31]. However, in remote homology-based predictions like this one, that aim at general function prediction and suggesting functional validation experiments, the purpose of actual structural model building is illustrative rather than predictive. Hence, water is not explicitly included.

The FAM69 proteins contain many cysteine residues (between 12 and 18 in human proteins), most of them conserved (see Fig. 2, left) and it has been suggested that most of them participate in disulphide bridges [20]. In our structure model of human FAM69A, some plausible S-S bridges can be postulated. One involves Cys293, next to the predicted active site aspartate D294, bridged to Cys331 (HCE motif located between strand  $\beta$ -9 and helix  $\alpha$ -F, in the region corresponding to the kinase activation loop) and possibly stabilising the predicted active site conformation



**Figure 1. CLANS graphs visualizing sequence similarities between protein kinase-like families.** Nodes represent sequences, edges represent similarity relationships. PSI-BLAST-detected significant (dark grey) and sub-significant (light grey) similarities shown. Dark green: kinase and kinase\_Tyr families, Red: FAM69; dark blue: SELO; yellow: alpha kinase; brown: UL97; cyan: PPKD; magenta: PI3\_P14; light green: PIP5K, orange: DUF1193 (FAM20); Black: other kinase families. The following P-value thresholds for significance of sequence similarity were used: 0.1 (top left), 0.001 (top right), 1E-5 (bottom left) and 1E-10 (bottom right). doi:10.1371/journal.pone.0066427.g001

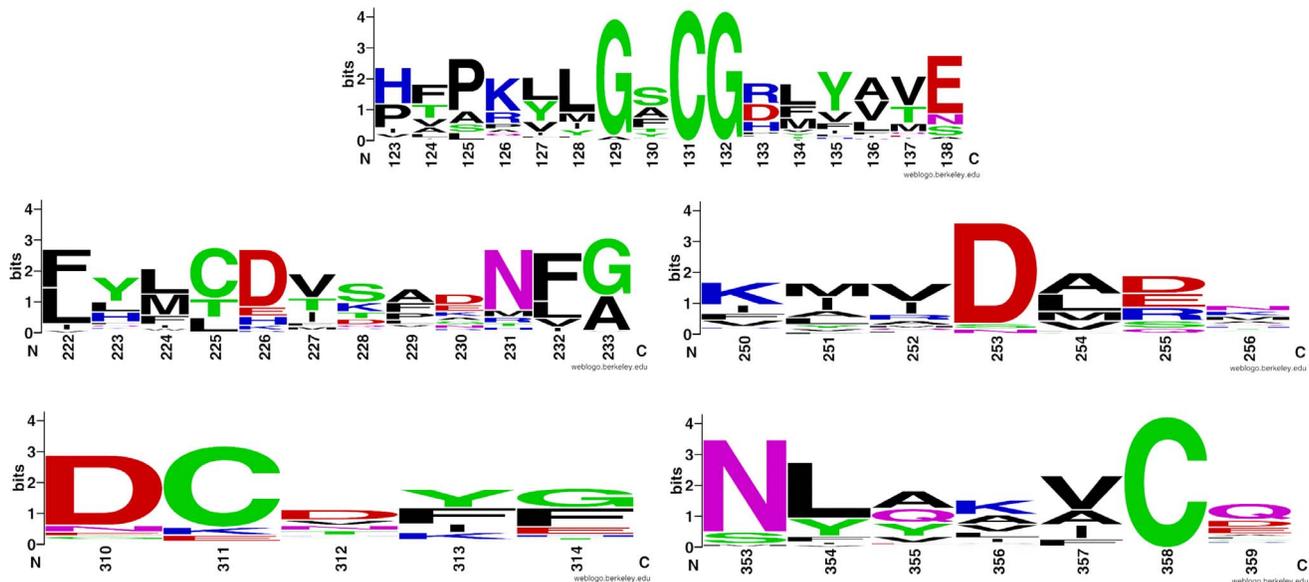
and positioning of D294. After manual adjustment of sidechain torsions of the two cysteine residues in the model, the sulphur-sulphur distance is 4.8 Å. Other possible disulphide bridges may involve six cysteine residues located within or near helices α-F, α-G and α-H and could stabilise the C-terminal lobe of FAM69s. In the FAM69A structure model, the sulphur atoms of Cys243 and Cys293 (from the DCR and SCI motives, located in helices α-F and α-H, respectively) are located within 3.19 Å. Further probable disulphide bridges could stabilise the N-terminal lobe and the ATP-binding region. Of note, some of cysteine residues conserved

in the FAM69 family are also found in PKDCC and SGK196 kinases (see Fig. 2). The cysteine residue of helix α-G is also conserved in many other known kinases, including the 4hcua, 3sxsA and 3s95A structures (see Fig. 2, left).

Nevertheless, the cysteine-rich kinase domain of FAM69s seems to be a unique arrangement among the known kinases. In the 1354 representatives of 27 kinase and putative kinase families as shown in Fig. 1, family-wide averages of cysteine residue count per kinase domain ranged from 0.6 to 10.6. Standard Analysis of Variation (ANOVA) test shows that the FAM69 family stands out as



**Figure 2. Multiple sequence alignments of selected FAM69 proteins.** Left: alignment of the kinase domain, covering the region 83–422 of FAM69A. The sequences are aligned using Promals3D (see Methods). Secondary structure prediction for human shown for selected proteins, for the solved structures actual secondary structure shown. Secondary structure elements named as in PKA [87]. Locations of predicted key catalytic residues shown, in standard PKA numbering (e.g. D166), as well as the ATP-binding loop (GxGxG). SwissProt identifiers shown for human sequences, otherwise, NCBI GI identifiers shown together with abbreviations of species names: Sp: sea urchin *Strongylocentrotus purpuratus*, Nv: sea cucumber *Nematostella vectens*, Dm: fruit fly *Drosophila melanogaster*; Mm: *Mus musculus*; Ce: *Caenorhabditis elegans*, Co: *Capsaspora owczarzakii*; S sp: *Salpingoeca sp. ATCC 50818*. Also shown selected close kinase homologues (PKDCC and SGK196) as well as sequences of selected kinases of known structures. Numbers in brackets indicate numbers of residues omitted from the alignment (shown only for the 1 cdk, 3 sxs structures, and for human FAM69A, DIA1, PKDCC and SGK196 sequences). R and C characters on black background above the alignment indicate the regulatory and catalytic spine residues, respectively [22]. The location of the the EF-hand motif shown, the motif itself is excised from the alignment and shown on the right. Right: alignment of the EF-hand region (corresponding to the region 165–199 of human FAM69A). Also shown EF-hand regions of human calmodulin and the 2PMY region used for model building. doi:10.1371/journal.pone.0066427.g002



**Figure 3. Kinase motifs.** Sequence logos for selected kinase motifs in the FAM69 family. *Top*: the  $\beta$ -4 -  $\beta$ -5 region. *Middle row, left*: the predicted active site region (corresponding to D166 and N171 in PKA). *Middle row, right*: the  $Mg^{2+}$  - binding motif (corresponding to the DFG motif in PKA). *Bottom, left*: predicted helix  $\alpha$ -F. *Bottom, right*: predicted helix  $\alpha$ -G. doi:10.1371/journal.pone.0066427.g003

significantly different from most other kinase families when kinase domains are compared for cysteine content (ANOVA p-value is less than  $10^{-4}$ ). FAM69 forms a homogenous group together with the viral UL97 family [32,33]. These viral kinases function intracellularly, in the nucleus and in the cytoplasm, facilitating viral infection by interfering with cellular processes such as cell cycle regulation and DNA replication. Of interest, the UL97 kinases have unusual optimal conditions for their catalytic activity (1.5 M NaCl and pH 9.5) [33]. Their high cysteine content may be related to these atypical preferred catalytic conditions. A kinase family following next, when sorted by cysteine count, is the DUF1193 (FAM20) [13,14]. This agrees with both FAM20 and FAM69 being known or predicted, respectively, to be extracellular. It has been accepted for a long time that, in general, extracellular proteins have higher cysteine and cystine content than intracellular ones [34–36].

When the human kinome alone [1] is analysed for cysteine content, again the FAM69 family stands out. In the chart (see Fig. 4), the FAM69 predicted kinase domains are found at the far right tail of cysteine count distribution. The difference between FAM69 and the established kinome is even more pronounced, because cysteine counts in kinase domains are imprecise due to imprecise definitions of domain boundaries. For example, some of the kinase domains with most cysteine counted (e.g. CDC7 and SRPK1) in fact possess Cys-rich insert regions within the kinase domain.

In contrast to FAM69, some of the known kinases that are termed cysteine-rich [37–39] are actually ones that contain specific cysteine-rich domains next to their kinase domains. In contrast, in the FAM69 family, the numerous cysteine residues are located within the kinase domain itself.

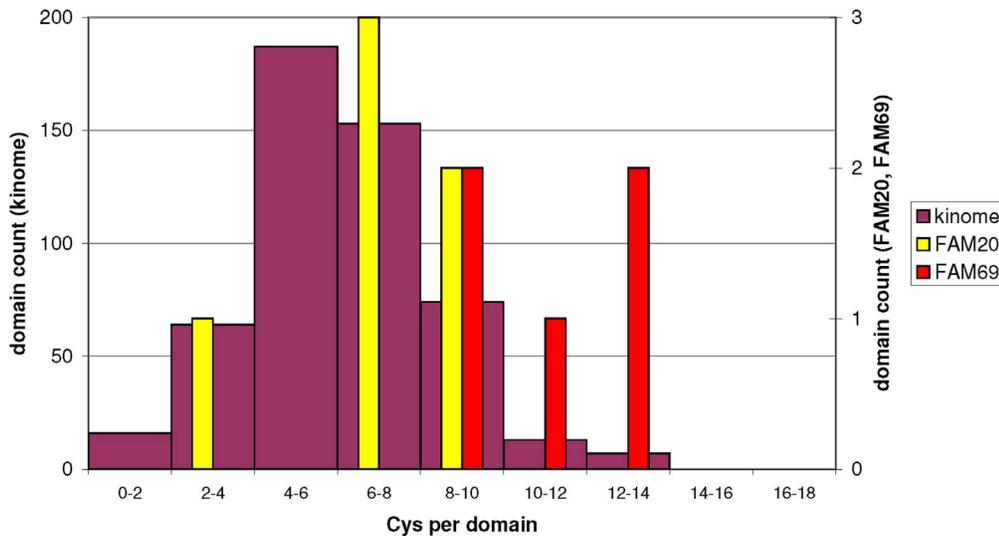
### Some FAM69s Contain a $Ca^{2+}$ -Binding Motif

Close to the predicted kinase ATP-binding loop in human FAM69A, a single EF-hand  $Ca^{2+}$ -binding motif can be detected by the HHpred algorithm, with p-value  $1E-11$ , in the region 165–190, approximately. The motif, inserted within the N-terminal

lobe of the kinase domain, most likely between helix C and strand  $\beta$ -4, is easily detected only in some FAM69s, e.g. human FAM69A and FAM69B, as well as one *Nematostella vectensis* protein, gi: 156376825, see Fig. 2 (right). In these proteins, the motif features all the residues necessary for the  $Ca^{2+}$ -binding activity, i.e.  $DxDxDGx[IV]xxxE$  in the ion-binding loop [40]. In other FAM69 sequences, including the other human sequences (FAM69C, C3ORF58 and CXORF36), some features of the EF-hand are visible, albeit due to many substitutions they are unlikely to bind calcium ions (See Fig. 2, right). To our knowledge, FAM69 is the first case of a protein-kinase like domain fused with an EF-hand motif in *Metazoa*. However, similar domain combinations are known in plants and some protists [41–43]. Yet, there, the EF-hands are located next to the kinase domains, not inserted into them. In humans, many protein kinases are regulated by calcium, but either by interaction with independent calcium sensors, e.g. calmodulin or by utilising specialised calcium-binding domains, C2, unrelated to EF-hands [44]. The unique location of the EF-hand domain within the FAM69 kinase domain, between helix  $\alpha$ -C and strand  $\beta$ -4 and near the predicted ATP site and predicted active site (see Fig. 5), suggests a regulatory role. This is supported by very good evolutionary conservation of the EF-hand domain in members of the FAM69 family from vertebrates and distant Metazoa, e.g. *Nematostella*. Since EF-hand motifs undergo  $Ca^{2+}$ -dependent dimerisation, one may speculate that FAM69A and FAM69B kinase activity depends on  $Ca^{2+}$ -binding-induced dimerisation.

### FAM69s are Implicated in Neurological Disorders

According to the PolyPhobius algorithm, FAM69A, B and C proteins possess transmembrane regions between residues 20–50, while C3ORF58 and CXORF36 have signal peptides (See Fig S3). Very similar predictions are obtained using the TMHMM tool. The TMHMM algorithm predicts extracellular location for all the five human FAM69 proteins (except the short cytoplasmic N-termini of FAM69A, B and C). The FAM69A, B and C proteins have been reported to localise to endoplasmic reticulum (ER) [20]



**Figure 4. Cysteines in kinases.** Histogram of cysteine residue count in kinase domains. Left scale, magenta bars: 516 kinase domains of the human kinome [1], Right scale, red bars: human FAM69 kinase domains; right scale, yellow bars: human FAM20 kinase domains. doi:10.1371/journal.pone.0066427.g004

as putative membrane-anchored molecules, while C3ORF58 (DIA1, GoPro49) has been shown to reside in the Golgi [45]. Tissue-wise, expression of FAM69A is ubiquitous while FAM69B and FAM69C are expressed mostly in the brain, the latter also in the eye [20].

C3ORF58 expression was observed in cartilaginous mesenchymal tissues, regulated developmentally, with highest expression seen in proliferating chondrocytes [45]. Further, colocalization with beta-coatamer protein was seen, suggestive of a function in membrane traffic [46]. Then, characteristic expression of C3ORF58 was observed in dental follicles, again suggestive of a role in trafficking and secretion [46]. The *CXORF36* gene has an ubiquitous expression pattern. The expression patterns of FAM69 genes and proteins, including brain, dental follicles, developing mesenchyma and cartilaginous cells, could be reconciled if one assumed participation in biological processes where substantial secretory activity is essential.

Consistently with brain-specific or brain-including expression pattern, several FAM69 genes were implicated in a number of neural disorders. One of two largest chromosome region deletions in autism involves the *C3ORF58* (*DIA1*) gene [47]. Of note, *C3ORF58* is up-regulated by neuronal activity, as shown by MEF2 RNAi assay [47].

The *CXORF36* (*DIA1R*) gene has been linked to the fragile X syndrome (FXS), with non-synonymous mutations found in this gene in two studies: S24P, K128R [48,49]. The molecular mechanisms underlying FXS are overlapping with those responsible for autism, since 30% of FXS patients develop autism [50]. In several publications, the Xp11.3 region that includes *CXORF36* has been linked to neurological disorders [51], including X-linked mental retardation (XLMR). Further, a gene in Xp11.3-4 region may contribute to the higher autism susceptibility in men [52]. Finally, recently, deletion of *CXORF36* was observed in the Kabuki syndrome, a congenital mental retardation syndrome [53].

The *FAM69A* gene has been linked to schizophrenia and bipolar disorder, with two intronic significant SNPs identified in a meta-analysis [54]. Also, the *FAM69A* region is the risk locus for multiple sclerosis, although other genes in that region may be the primary culprits [55].

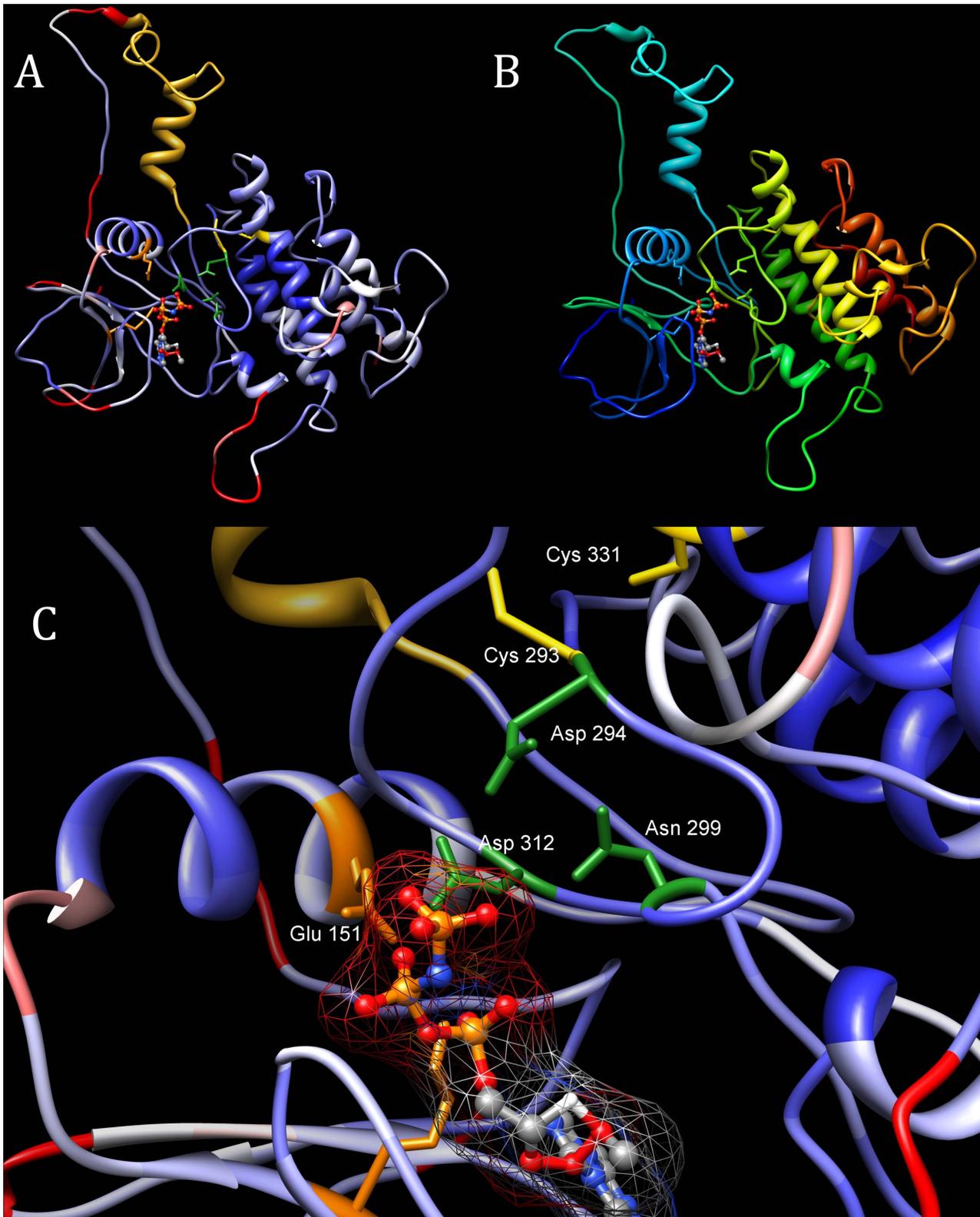
An analysis of rare copy number variation (CNV) in autism spectrum disorders found variation in three FAM69 genes: *FAM69B*, *C3ORF58*, *CXORF36*-2028680287 [56]. Also, a deletion of *FAM69B* in autism has been observed [57]. A network-based analysis of genes with CNV in autism identified involvement of synapse formation and function processes [58].

## Discussion

We present the discovery of a novel putative kinase family with members in humans and presence throughout *Metazoa* as a yet another small step towards filling in the blank spots in the complex regulatory machinery of the living cell. Charting of the kinome is important for unbiased advancement of biology and medicine [9,59].

Can the kinase function prediction for FAM69s be trusted, or is it only a reliable three-dimensional fold prediction? Conservation of key residues and evolutionary conservation in *Metazoa* suggest indeed a conserved kinase function. For very distant homologues, the sequence alignment details are known to be less reliable than the overall detection of homology [60]. Thus, some of our definitions of FAM69 predicted active site motifs (e.g. location of the residue corresponding to E91 of PKA) or secondary structure element assignments may be erroneous. Also, it is not straightforward to predict the substrate, yet, FAM69 similarity to classic protein kinases suggests FAM69 proteins are kinases that phosphorylate proteins. An exception is one of the FAM69 proteins, *CXORF36* (*DIA1R*), a protein restricted to vertebrates (see FigS1) [61]. Although it clearly is a homologue of C3ORF58 (*DIA1*), it does not have the predicted active site aspartate conserved (corresponding to D166 in PKA). Thus, *CXORF36* is probably a pseudokinase that may interfere with signalling by other FAM69 proteins in a dominant negative fashion. Alternatively, *CXORF36* may be a highly atypical kinase.

Precise placement of FAM69s in the protein kinase clan is difficult due to the presence of the EF-hand insert and sequence divergence (see Fig. 1). Closest known kinase homologues seem to be the PKDCC and the SGK196 subfamilies of the pkinase and pkinase\_Tyr families (see Fig. S4). The uncharacterised SGK196 kinase group is present in Opisthokonts (*Metazoa*, *Choanoflagellates*,



**Figure 5. Structure model of the kinase domain of human FAM69A.** *Top left:* model coloured by MetaMQAP model quality score (blue: good quality, red: poor quality). On left, the EF-hand motif is shown in yellow. *Top right:* as in Fig. 5 (top left), model coloured by sequence: from dark blue (N-terminus) to dark red (C-terminus). *Bottom:* close-up of the predicted active site with ATP molecule bound. Side chains of key predicted active site residues shown: D294 (PKA numbering: 166), N299 (171), D312 (184), also the two cysteines near the predicted active site that may form a S-S bridge: C293 and C331.

doi:10.1371/journal.pone.0066427.g005

and the early-branching opisthokont *Capsaspora*, but not *Fungi*). It also has remote homologues in plants. Of note, the atypical MCD motif found in the predicted active site of FAM69A instead of [HY]RD (the catalytic motif of typical kinases), is present also in most SGK196 homologues, including a primitive unicellular opisthokont, filasterean *Capsaspora owczarzewski* [62,63], a choanoflagellate, *Salpingoeca* sp. ATCC 50818 [64], and one of simplest multicellular animals, the plocozoan *Trichoplax*. Similarly to FAM69A, B and C, the SGK196 protein is predicted to be an extracellular protein possessing an N-terminal transmembrane segment. The second protein group clearly homologous to FAM69, is the PKDCC subfamily (protein kinase domain containing, cytoplasmic; SgK493; Vertebrate lonesome kinase, VLK [65]). Contrary to the VLK designation (vertebrate-specific), PKDCC is present in Metazoa (e.g. *Nematostella*, *Strongylocentrotus*). PKDCC is enriched in the Golgi apparatus and regulates transport from the Golgi to the plasma membrane [66]. On a systemic level, PKDCC regulates lung and bone development [65,66]. Lack of PKDCC results in morphological abnormalities, such as linked to deficient biomineralisation [65]. Recently, other Golgi-localised kinases have gained interest, namely the recently discovered Metazoan FAM20 kinases (FAM20C, FAM20B and Four-Jointed) that are not closely related to FAM69s by sequence similarity. These kinases have been demonstrated to phosphorylate secreted proteins (or xylose in the case of FAM20B). FAM20B and Four-Jointed reside in the Golgi while FAM20C has been observed both in the Golgi and extracellularly.

Taking together the information on FAM69 predicted kinases, it is all too tempting to speculate that indeed, FAM69s carry out functions similar to the PKDCC and FAM20 novel kinases [11–14,66]. Thus, different FAM69 proteins may be involved in phosphorylation of secreted proteins, or in regulating the transport from ER to Golgi and from Golgi to the plasma membrane. Thus, one may speculate that some of FAM69s may embody the yet unidentified kinases regulating of ER-to-Golgi vesicle transport [67,2011,H89]. It is recognised that vesicular trafficking is critical in neuron development and its malfunctions may result in mental retardation [68]. Thus, the neurological disorders related to FAM69 genes may have a common denominator, malfunction of the secretory pathway in neurons [19].

Evolutionary origin of FAM69 predicted kinases is obviously within the protein kinase-like (PKL) clan. In contrast to FAM69s, PKDCC and SGK196 clearly belong to the classic pkinase and pkinase\_Tyr families, respectively (Pfam database identifiers PF00069, PF07714), according to the HMMER tool (see Fig. S4). Because the phylogenetic spread of the SGK196 subfamily is broader than the spread of PKDCC subfamily of the FAM69s, one may speculate that the FAM69 family originated from a SGK196-like ancestor early in Metazoan evolution. It has been noted previously that kinase repertoire essential for multicellular life originated in pre-Metazoan unicellular Eukaryotes [69].

Translating a structure prediction into a useful function prediction is a challenge. Here, we strove to achieve this, complementing structure predictions with analyses of available functional data and literature. Yet, the ultimate answers as to the functions of the FAM69 family will only come from experiments.

## Methods

For remote homology identification, PSI-BLAST searches used the standard parameters on nr database at NCBI as of 09.2012. For domain assignments, HMMER3 [70] on the Pfam database as of 09.2012 was used.

For survey of similarities within the kinase-like clan (Fig. 1), the CLANS algorithm [18] was run on a set of sequences including a) all Pfam seeds from the 17 families of the protein kinase-like clan (CL0016), b) the seeds from FAM69 family (PIP49\_C, Pfam:PF12260), c) representative SELO domains [10], d) seeds from the Pfam families: Alpha\_kinase (PF02816), PI3\_PI4\_kinase (PF00454), Act-Frag\_cataly\_kinase, PF09192, PPKD\_N (PF01326), PIP5K (PF01504), e) the DUF1194 family (FAM20) [12,14]. For the d) group, structural similarity to the PKL kinases is known [3]. CLANS was run with 5 iterations of PSI-BLAST, using the BLOSUM45 substitution matrix and inclusion threshold 0.001. For the CLANS graphs, sequence similarity relations with significance of P-values below 0.1, 0.001, 1E-5 and 1E-10 were considered, as indicated in the Figure 1.

For closer examination of similarities between FAM69 and other kinases (Fig. S4), a set of representative sequences was built as follows. The kinase domain regions of the human FAM69A, DIA1, PKDCC and SGK196 proteins (as outlined in Fig. 2, left) were used to find representative homologous sequences by running one iteration of JackHMMER [71] on the RefSeqP database with 1E-5 E-value threshold. The representative sequences found were cleared of redundancy above 70% sequence identity using CD-HIT [72]. Then, the sequences obtained plus seeds for the pkinase and pkinase\_Tyr Pfam families were input to the CLANS algorithm, run on nr90 and env\_nr90 databases with the Blosum45 matrix and 3 iterations of PSI-Blast. The cutoff for inclusion of a relation in the CLANS graph building was P-value equal to 1E-5.

Transmembrane region predictions were achieved by the TMHMM and Phobius servers [73,74]. The Jpred and PsiPred servers were used to predict the secondary structures [75,76]. Multiple sequence alignments of FAM69A, B, C and DIA1, DIA1R subfamilies were built using the MUSCLE program [77]. The final multiple sequence alignment was built using Promals3D, with additional alignment constraints for selected kinase motifs [78]. The constraints were derived from HHpred alignments of human FAM69A, DIA1, PKDCC and SGK196 to the 3sxsA sequence. HHpred was run using MUSCLE-generated alignments of close homologues of the above-listed four proteins.

The structure of human FAM69A was modelled by comparative modelling. The templates were chosen based on results of FFAS03 predictions run on PDB database. The FFAS03 method [15] that uses sequence profile-to-profile comparison was supplemented by the HHpred algorithm [79] that employs HMM-to-HMM comparison. Best scoring kinase structures (FFAS03 score below -20 and percent identity above 9) were picked up for further analysis. The modelled (target) structure was constructed based on pairwise alignments of FAM69A sequence and the two templates identified by FFAS03 (3HGK\_A: kinase Pto from *Solanum pimpinellifolium* [80] and 3S95\_A: human LIMK1 kinase domain [81]). Since analysis of FAM69A sequence indicated the presence of the EF-hand Ca<sup>2+</sup>-binding motif, which is absent in typical kinases, the model was constructed using combined templates, incorporating an EF-hand domain into the template. According to FFAS03 score, the 2PMY structure (EF-domain of human RASEF, region 59–89) was selected for modelling purposes. Because the selected templates do not contain ATP ligands, the structure of protein kinase A complexed with ANP and Mn<sup>2+</sup> (PDB:1CDK) was used to identify the possible ANP and Mn<sup>2+</sup> ligand positions in 3HGK structure and to manually place ANP structure in the model. The two kinase templates were aligned using FATCAT (Flexible structure alignment by chaining aligned fragment pairs allowing twists) [82] and the resultant alignment served for modelling. We combined FATCAT,

HHpred and FFAS03 pairwise alignments to refine the multiple alignment between the modelled sequence and selected templates. The alignment was also manually adjusted to accommodate predicted secondary structures (See Fig. S2).

The structure model was constructed automatically by the program MODELLER9 v 8 [83] using the combined scripts: homology modelling with multiple templates and model-ligand.

The MetaMQAP server [84] was used to estimate the correctness of the 3D models using a number of model quality assessment methods in a meta-analysis.

For sequence logos, the WebLogo tool was used [85]. The sequences represented in the logo are obtained by the JackHMMer tool [71] using the sequence alignment from Fig. 2 (left) as query against the RefSeq database, and similarity threshold at bitscore 70 to avoid distant homologues.

For presentation of multiple sequence alignments, the BioEdit software was used [86].

## Supporting Information

**Figure S1 Phylogenetic tree of the FAM69 sequences obtained using the alignment from Fig. 2 (left) and the PhyML maximum likelihood algorithm [88]. Approximate bootstrap values shown.**

(PDF)

**Figure S2 Target/template alignment used for structure modeling.**

(PDF)

**Figure S3 Transmembrane helix and signal peptide predictions for human FAM69 proteins, obtained by the Phobius algorithm.**

(PDF)

**Figure S4 CLANS analysis for FAM69, SGK196 and PKDCC proteins together with pkinase and pkinase\_Tyr families.** Dark blue: FAM69ABC subfamily, light blue: DIA1 subfamily. Dark green: PKDCC homologues. Light green: SGK196 homologues. Orange: pkinase\_Tyr family seeds, Red: pkinase family seeds. Sequence similarity relations with significance of P-value below 1E- considered.

(TIF)

## Author Contributions

Conceived and designed the experiments: KP. Performed the experiments: KP MD AL. Analyzed the data: KP MD AL. Contributed reagents/materials/analysis tools: KP MD. Wrote the paper: KP.

## References

- Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. *Science* 298: 1912–34.
- Kamman N, Taylor SS, Zhai Y, Venter JC, Manning G (2007) Structural and functional diversity of the microbial kinome. *PLoS Biol* 5: e17.
- Cheek S, Ginalski K, Zhang H, Grishin NV (2005) A comprehensive update of the sequence and structure classification of kinases. *BMC Struct Biol* 5: 6.
- Taylor SS, Radzio-Andzelm E (1994) Three protein kinase structures define a common motif. *Structure* 2: 345–55.
- Manning G, Plowman GD, Hunter T, Sudarsanam S (2002) Evolution of protein kinase signaling from yeast to man. *Trends Biochem Sci* 27: 514–20.
- Eglen R, Reisine T (2011) Drug discovery and the human kinome: Recent trends. *Pharmacol Ther* 130: 144–56.
- Manning BD (2009) Challenges and opportunities in defining the essential cancer kinome. *Sci Signal* 2: pe15.
- Fedorov O, Muller S, Knapp S (2010) The (un)targeted cancer kinome. *Nat Chem Biol* 6: 166–169.
- Edwards AM, Isserlin R, Bader GD, Frye SV, Willson TM, et al. (2011) Too many roads not taken. *Nature* 470: 163–5.
- Dudkiewicz M, Szczepinska T, Grynberg M, Pawlowski K (2012) A novel protein kinase-like domain in a selenoprotein, widespread in the tree of life. *PLoS One* 7: e32138.
- Ishikawa HO, Takeuchi H, Haltiwanger RS, Irvine KD (2008) Four-jointed is a Golgi kinase that phosphorylates a subset of cadherin domains. *Science* 321: 401–4.
- Koike T, Izumikawa T, Tamura J, Kitagawa H (2009) FAM20B is a kinase that phosphorylates xylose in the glycosaminoglycan-protein linkage region. *Biochem J* 421: 157–62.
- Ishikawa HO, Xu A, Ogura E, Manning G, Irvine KD (2012) The Raine Syndrome Protein FAM20C Is a Golgi Kinase That Phosphorylates Bio-Mineralization Proteins. *PLoS One* 7: e42988.
- Tagliabracchi VS, Engel JL, Wen J, Wiley SE, Worby CA, et al. (2012) Secreted Kinase Phosphorylates Extracellular Proteins that Regulate Biomineralization. *Science* 336: 1150–3.
- Rychlewski L, Jaroszewski L, Li W, Godzik A (2000) Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci* 9: 232–41.
- Samir AA, Ropolo A, Grasso D, Tomasini R, Dagorn JC, et al. (2000) Cloning and expression of the mouse PIP49 (Pancreatitis Induced Protein 49) mRNA which encodes a new putative transmembrane protein activated in the pancreas with acute pancreatitis. *Mol Cell Biol Res Commun* 4: 188–93.
- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389–402.
- Frickey T, Lupas A (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics* 20: 3702–4.
- Aziz A, Harrop SP, Bishop NE (2011) Characterization of the deleted in autism 1 protein family: implications for studying cognitive disorders. *PLoS One* 6: e14547.
- Tennant-Eyles AJ, Moffitt H, Whitehouse CA, Roberts RG (2011) Characterisation of the FAM69 family of cysteine-rich endoplasmic reticulum proteins. *Biochem Biophys Res Commun* 406: 471–7.
- Hanks SK, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* 241: 42–52.
- Taylor SS, Kornev AP (2012) Protein kinases: evolution of dynamic regulatory proteins. *Trends Biochem Sci* 36: 65–77.
- Gooneskere NC, Shipley K, O'Connor K (2010) The challenge of annotating protein sequences: The tale of eight domains of unknown function in Pfam. *Comput Biol Chem* 34: 210–4.
- Jaroszewski L, Li Z, Krishna SS, Bakolitsa C, Wooley J, et al. (2009) Exploration of uncharted regions of the protein universe. *PLoS Biol* 7: e1000205.
- Bateman A, Coggill P, Finn RD (2010) DUFs: families in search of function. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 66: 1148–52.
- Pawlowski K, Lepisto M, Meinander N, Sivars U, Varga M, et al. (2006) Novel conserved hydrolase domain in the CLCA family of alleged calcium-activated chloride channels. *Proteins-Structure Function and Bioinformatics* 63: 424–439.
- Yurtsever Z, Sala-Rabanal M, Randolph DT, Scheaffer SM, Roswit WT, et al. (2012) Self-cleavage of Human CLCA1 Protein by a Novel Internal Metalloprotease Domain Controls Calcium-activated Chloride Channel Activation. *J Biol Chem* 287: 42138–49.
- Pawlowski K, Pio F, Chu Z, Reed JC, Godzik A (2001) PAAD - a new protein domain associated with apoptosis, cancer and autoimmune diseases. *Trends Biochem Sci* 26: 85–7.
- Bonardi V, Cherkis K, Nishimura MT, Dangl JL (2012) A new eye on NLR proteins: focused on clarity or diffused by complexity? *Curr Opin Immunol* 24: 41–50.
- Papioian GA, Ulander J, Eastwood MP, Luthey-Schulten Z, Wolynes PG (2004) Water in protein structure prediction. *Proc Natl Acad Sci U S A* 101: 3352–7.
- Kaszuba K, Róg T, Bryl K, Vattulainen I, Karttunen M (2010) Molecular Dynamics Simulations Reveal Fundamental Role of Water As Factor Determining Affinity of Binding of  $\beta$ -Blocker Nebivolol to  $\beta$ 2-Adrenergic Receptor. *The Journal of Physical Chemistry B* 114.
- Prichard MN (2009) Function of human cytomegalovirus UL97 kinase in viral infection and its inhibition by maribavir. *Rev Med Virol* 19: 215–29.
- Gershburg E, Pagano JS (2008) Conserved herpesvirus protein kinases. *Biochim Biophys Acta* 1784: 203–12.
- Fahy RC, Hunt JS, Windham GC (1977) On the cysteine and cystine content of proteins. Differences between intracellular and extracellular proteins. *J Mol Evol* 10: 155–60.
- Beeby M, O'Connor BD, Ryttersgaard C, Boutz DR, Perry LJ, et al. (2005) The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol* 3: e309.
- Feige MJ, Hendershot LM (2011) Disulfide bonds in ER protein folding and homeostasis. *Curr Opin Cell Biol* 23: 167–75.
- Stiegler AL, Burden SJ, Hubbard SR (2009) Crystal structure of the frizzled-like cysteine-rich domain of the receptor tyrosine kinase MuSK. *J Mol Biol* 393: 1–9.

38. Molendijk AJ, Ruperti B, Singh MK, Dovzhenko A, Ditungou FA, et al. (2008) A cysteine-rich receptor-like kinase NCRK and a pathogen-induced protein kinase RBK1 are Rop GTPase interactors. *Plant J* 53: 909–23.
39. Hommel U, Zurini M, Luyten M (1994) Solution structure of a cysteine rich domain of rat protein kinase C. *Nat Struct Biol* 1: 383–7.
40. Grabarek Z (2006) Structural basis for diversity of the EF-hand calcium-binding proteins. *J Mol Biol* 359: 509–25.
41. Chandran V, Stollar EJ, Lindorff-Larsen K, Harper JF, Chazin WJ, et al. (2006) Structure of the regulatory apparatus of a calcium-dependent protein kinase (CDPK): a novel mode of calmodulin-target recognition. *J Mol Biol* 357: 400–10.
42. DeFalco TA, Bender KW, Snedden WA (2010) Breaking the code: Ca<sup>2+</sup> sensors in plant signalling. *Biochem J* 425: 27–40.
43. Boudsocq M, Droillard MJ, Regad L, Lauriere C (2012) Characterization of Arabidopsis calcium-dependent protein kinases: activated or not by calcium? *Biochem J* 447: 291–9.
44. Farah CA, Sossin WS (2012) The role of C2 domains in PKC signaling. *Adv Exp Med Biol* 740: 663–83.
45. Takatalo M, Jarvinen E, Laitinen S, Thesleff I, Ronnholm R (2008) Expression of the novel Golgi protein GoPro49 is developmentally regulated during mesenchymal differentiation. *Dev Dyn* 237: 2243–55.
46. Takatalo MS, Tummers M, Thesleff I, Ronnholm R (2009) Novel Golgi protein, GoPro49, is a specific dental follicle marker. *J Dent Res* 88: 534–8.
47. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, et al. (2008) Identifying autism loci and genes by tracing recent shared ancestry. *Science* 321: 218–23.
48. Jensen LR, Lenzner S, Moser B, Freude K, Tzschach A, et al. (2007) X-linked mental retardation: a comprehensive molecular screen of 47 candidate genes from a 7.4 Mb interval in Xp11. *Eur J Hum Genet* 15: 68–75.
49. Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, et al. (2009) A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet* 41: 535–43.
50. Hagerman R, Hoem G, Hagerman P (2010) Fragile X and autism: Intertwined at the molecular level leading to targeted treatments. *Mol Autism* 1: 12.
51. Thiselton DL, McDowall J, Brandau O, Ramser J, d'Esposito F, et al. (2002) An integrated, functionally annotated gene map of the DXS8026-ELK1 interval on human Xp11.3–Xp11.23: potential hotspot for neurogenetic disorders. *Genomics* 79: 560–72.
52. Good CD, Lawrence K, Thomas NS, Price CJ, Ashburner J, et al. (2003) Dosage-sensitive X-linked locus influences the development of amygdala and orbitofrontal cortex, and fear recognition in humans. *Brain* 126: 2431–46.
53. Lederer D, Grisart B, Digilio MC, Benoit V, Crespin M, et al. (2012) Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. *Am J Hum Genet* 90: 119–24.
54. Wang KS, Liu XF, Aragao N (2010) A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr Res* 124: 192–9.
55. Alcina A, Fernandez O, Gonzalez JR, Catala-Rabasa A, Fedetz M, et al. (2010) Tag-SNP analysis of the GF11-EVI5-RPL5-FAM69 risk locus for multiple sclerosis. *Eur J Hum Genet* 18: 827–31.
56. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, et al. (2010) Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 466: 368–72.
57. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, et al. (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* 70: 863–85.
58. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, et al. (2011) Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses. *Neuron* 70: 898–907.
59. Hanson AD, Pribat A, Waller JC, de Crecy-Lagard V (2010) 'Unknown' proteins and 'orphan' enzymes: the missing half of the engineering parts list—and how to find it. *Biochem J* 425: 1–11.
60. Jaroszewski L, Rychlewski L, Godzik A (2000) Improving the quality of twilight-zone alignments. *Protein Sci* 9: 1487–96.
61. Aziz A, Harrop SP, Bishop NE (2011) DIA1R is an X-linked gene related to Deleted In Autism-1. *PLoS One* 6: e14534.
62. Ruiz-Trillo I, Inagaki Y, Davis LA, Sperstad S, Landfald B, et al. (2004) Capsaspora owczarzaki is an independent opisthokont lineage. *Curr Biol* 14: R946–7.
63. Shalchian-Tabrizi K, Minge MA, Espelund M, Orr R, Ruden T, et al. (2008) Multigene phylogeny of choanozoa and the origin of animals. *PLoS One* 3: e2098.
64. Torruella G, Derelle R, Paps J, Lang BF, Roger AJ, et al. (2012) Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol Biol Evol* 29: 531–44.
65. Imuta Y, Nishioka N, Kiyonari H, Sasaki H (2009) Short limbs, cleft palate, and delayed formation of flat proliferative chondrocytes in mice with targeted disruption of a putative protein kinase gene, *Pkccc* (AW548124). *Dev Dyn* 238: 210–22.
66. Kinoshita M, Era T, Jakt LM, Nishikawa S (2009) The novel protein kinase Vlk is essential for stromal function of mesenchymal cells. *Development* 136: 2069–79.
67. Nakagawa H, Miyazaki S, Abe T, Umadome H, Tanaka K, et al. (2011) H89 sensitive kinase regulates the translocation of Sar1 onto the ER membrane through phosphorylation of ER-coupled beta-tubulin. *Int J Biochem Cell Biol* 43: 423–30.
68. Giannandrea M, Bianchi V, Mignogna ML, Sirri A, Carrabino S, et al. (2010) Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly. *Am J Hum Genet* 86: 185–95.
69. Suga H, Dacre M, de Mendoza A, Shalchian-Tabrizi K, Manning G, et al. (2012) Genomic survey of premetazoans shows deep conservation of cytoplasmic tyrosine kinases and multiple radiations of receptor tyrosine kinases. *Sci Signal* 5: ra35.
70. Eddy SR (2011) Accelerated Profile HMM Searches. *PLoS Comput Biol* 7: e1002195.
71. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39: W29–37.
72. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–9.
73. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol* 6: 175–82.
74. Kall L, Krogh A, Sonnhammer EL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 35: W429–32.
75. Cole C, Barber JD, Barton GJ (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36: W197–201.
76. McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16: 404–5.
77. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
78. Pei J, Kim BH, Grishin NV (2008) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Res* 36: 2295–300.
79. Soding J, Biegert A, Lupas AN (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33: W244–8.
80. Dong J, Xiao F, Fan F, Gu L, Cang H, et al. (2009) Crystal structure of the complex between *Pseudomonas* effector AvrPtoB and the tomato Pto kinase reveals both a shared and a unique interface compared with AvrPto-Pto. *Plant Cell* 21: 1846–59.
81. Beltrami A, Chaikuad A, Daga N, Elkins JM, Mahajan P, et al. (2011) Crystal structure of the human LIMK1 kinase domain in complex with staurosporine; Available from: <http://pdb.rcsb.org/pdb/explore/explore.do?structureId=3s95>.
82. Ye Y, Godzik A (2003) Flexible structure alignment by chaining aligned fragment pairs allowing twists. *Bioinformatics* 19 Suppl 2: ii246–55.
83. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779–815.
84. Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM (2008) MetaMQAP: a meta-server for the quality assessment of protein models. *BMC Bioinformatics* 9: 403.
85. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14: 1188–90.
86. Hall T (2012) BioEdit. 2012; Available from: <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>.
87. Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, et al. (1991) Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* 253: 407–14.
88. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, et al. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36: W465–9.