A novel predicted Ca^{2+}-regulated kinase family implicated in neurological disorders

Małgorzata Dudkiewicz[2], Anna Lenart[1], and Krzysztof Pawłowski[1,2*]

Keywords: protein kinase, neurological disorder, autism, protein structure prediction, Ca2+-regulated, EF-hand

[1] Nencki Institute of Experimental Biology, PAS, Pasteura 3, 02-093 Warsaw, Poland

[2] Warsaw University of Life Sciences, Nowoursynowska 166, 02-787 Warsaw, Poland

[*] corresponding author:

e-mail: k.pawlowski@nencki.gov.pl

* - to whom correspondence should be addressed

Abstract

Protein kinases are essential effectors of cellular signaling. Surprisingly, using bioinformatics tools, we predicted protein kinase structure and function for proteins of unknown function (FAM69 family) coded by five related human genes and their Metazoan homologues. Analysis of three-dimensional structure models and conservation of the classic catalytic motifs of protein kinases in four of human FAM69 proteins suggests they might have retained catalytic phosphotransferase activity. The FAM69 genes, *FAM69A, FAM69B, FAM69C, C3ORF58* and *CXORF36,* are by large uncharacterized molecularly, yet linked to several neurological disorders in genetics studies. An EF-hand Ca^{2+}-binding domain in FAM69A and FAM69B proteins, inserted within the structure of the kinase domain, suggests they function as Ca^{2+}-dependent kinases.

Background

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 another common problem in biology. Here, we marry the two: a protein family with no function and function with no protein.

Protein kinase-like (PKL) proteins are a huge grouping of regulatory / signalling and biosynthetic enzymes [1, 2], regulating most processes in a living cell, 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 oveall more than

half million PubMed articles as of June 2011), the research effort has been biased[7, 8] whereas approx. 10% of known kinases yield at least 90% of publications [9]. Also, the kinome may be not fully charted yet, e.g. a protein kinase-like domain has been recently discovered in selenoprotein O (K. Pawłowski, in preparation).

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. First, we prove that FAM69 family members are homologous to protein kinases. Second, we predict that most of FAM69s do have protein kinase activity. Third, we predict that some of 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.

Results

FAM69s belong to protein kinase-like clan

Examination of protein families distantly similar to the classical protein kinases, according to results of the FFAS algorithm, suggested that the uncharacterised Pfam family PIP49_C (Pfam:PF12260, Pancreatitis induced protein 49 C terminal) [10] may be homologous to protein kinases. Starting from the human FAM69A protein sequence, [Swiss-Prot:FA69A_HUMAN], in the second PSI-Blast iteration [11] one obtains significant similarity to a human Ser/Thr protein kinase PKDCC [GenBank:292495024], E-value 4e-08, 21% sequence identity over 197 residues. 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 similarity covers the region 225-415 of the human FAM69A protein, that aligns to the C-terminal lobe of the kinase domain. CLANS clustering analysis, presenting in a graph

form distant relationships between groups of proteins, reflects relatively close similarity of FAM69s to the many known kinase families (Fig. 1).

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 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.

The FAM69 family is present in most Metazoan branches, with the sea anemone *Nematostella vectensis* being the organism most distant from humans possessing it. *Nematostella* has four FAM69 genes, three of which code proteins with the typical FAM69 features conserved. FAM69s can be clearly separated into two major branches, one containing human FAM69A, FAM69B, FAM69C and the other containing C3ORF58 and CXORF36 (not shown). The two protein groups have been analysed bioinformatically [12]; [13], but they have never been identified as homologues before, although their similarity is significant(FFAS Zscore -56 between human FAM69A and C3ORF56, see also Fig. 2).

Most FAM69s are probably 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 similarity to the C-terminal lobe of the classical kinase structure. 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). A sequence alignment ambiguity that does not affect the general conclusions arises from the uncertainty as to the precise locations of domain boundaries, a

problem well-known. Depending on the extent of the inserted EF-hand region, different regions of FAM69A align to the helix α-C of PKA, with the E91 of PKA corresponding to either E133 or E202 in FAM69A. The latter case seems more plausible since this residue is more conserved in the FAM69 family, also it is in a region consistently predicted to be alpha-helical.

Several of the conserved kinase regions, as defined for protein kinase A (PKA) [14], are clearly conserved in FAM69s (See Fig. 2). The presumed catalytic base (D166 in PKA) is present in most FAM69s, excluding CXORF36, albeit in a [LM]CD motif instead of [HY]RD. Also well-conserved are residues corresponding to N171 and D184 of PKA, responsible for binding the Mg^{2+} ions. Also conserved are K72 and E91 involved in binding the phosphate groups of the ATP molecule.

In cases of remote sequence similarity, building a structural model is of an illustrative nature, yet it serves also as a feasibility check for the predicted structure. We analysed two structural models of the FAM69 kinase domain (residues 79-404, excluding the EF-hand domain region) built using the archetypical kinase structure, PKA, as a template and two alternative alignments in the ambiguous region near helix α-C (See Fig. 3). For FAM69 structure model building, other kinase structures could have been used, possibly more similar. However, we used the classical PKA structure as a template since the main aim of model building was illustration and interpretation of the function prediction. The catalytic core of the protein kinases contains an unique ATP-binding motif. In FAM69s, structure prediction algorithms do not allow unambiguous detection of a corresponding site. However, an analysis of weak fold predictions, together with secondary structure predictions allowed detection of similarity to the kinase N-terminal lobe. Since the Gly-rich motif is not present in FAM69s in the predicted β-1-β-2 loop, it is likely that an atypical ATP-binding mode is employed in the FAM69 family, possibly using the conserved proline residue of that loop. Overall, both structure models are reasonable as judged by the MetaMQAP model quality scoring, with slightly more favourable scores for the model which places helix α-C in the region of Glu202, not that of Glu133, as also suggested by conservation analysis.

FAM69s contain many cysteine residues (between 12 and 18 in human proteins), and it has been suggested that most of them participate in disulphide bridges [13]. In the model, some plausible S-S bridges can be postulated. One involves C293, next to the active site D294, bridged to C331 (located between helices α-EF, α-F) and possibly stabilising the active site conformation and postitioning of D294. Other possible disulphide bridges may involve six cysteine residues located within or near helices α-F and α-G and could stabilise the C-terminal lobe of FAM69s. More probable disulphide bridges could stabilise the N-terminal lobe and the ATP-binding region. The cysteine-rich kinase domain seems to be a unique arrangement among the known kinases. For example, in 3055 kinase seed sequences of protein kinases in Pfam database, a kinase domain contains on average 3.75 Cys residues, compared to 11 in FAM69A.

Some FAM69s contain a Ca2+-binding motif

Close to the predicted kinase ATP-binding loop in human FAM69A, a single EF-hand Ca2+-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 β-4, is easily detected only in some FAM69s, e.g. human FAM69A and FAM69B as well as one *Nematostella vectensis* protein, [GenBank:156376825], see Fig. 2. 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 [15]. 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. To our knowledge, this is the first report 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. 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.

The unique location of the EF-hand domain within the kinase domain, and near the predicted ATP site and active site (see Fig. 3) strongly suggests a regulatory role.

FAM69s are involved in neurological disorders

According to the Phobius algorithm, FAM69A, B and C proteins possess transmembrane regions between residues 20-50, while C3ORF58 and CXORF36 have signal peptides. Similar predictions are obtained using the TMHMM tool. The three FAM69 proteins have been reported to localise to ER [13] as putative membrane-anchored molecules, while C3ORF58 (DIA1, LOC205428, GoPro49) has been shown to reside in the Golgi [16]. Tissue-wise expression of FAM69A is ubiquitous while FAM69B and FAM69C are expressed mostly in the brain, the latter also in the eye [13].

C3ORF58 expression was observed in cartilaginous mesenchymal tissues, regulated developmentally, with highest expression seen in proliferating chondrocytes [16]. Further, colocalization with beta-coatomer protein was seen, suggestive of a function in membrane traffic [17]. Then, characteristic expression of C3ORF58 was observed in dental follicles, suggestive of a role in trafficking and secretion [17]. The *CXORF36* gene has an ubiquitous expression pattern. The FAM69 expression patterns, 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 [18]. Of note, *C3ORF58* is up-regulated by neuronal activity, as shown by MEF2 RNAi assay [18].

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 [19, 20]. The molecular mechanisms underlying FXS are overlapping, with those responsible for autism, since 30% of FXS patients develop autism[21]. In several publications, the Xp11.3 region that includes *CXORF36* has been linked to neurological disorders [22], including X-linked mental retardation (XLMR). Further, a gene in Xp11.3-4 region may contribute to the higher autism susceptibility in men [23].

The *FAM69A* gene has been linked to schizophrenia and bipolar disorder, with two intronic significant SNPs identified in a meta-analysis [24]. Also, the *FAM69A* region is the risk locus for multiple sclerosis, although other genes in that region may be the primary culprits [25].

An analysis of rare copy number variation (CNV) in autism spectrum disorders found variation in three FAM69 genes: *FAM69B, C3ORF58, CXORF36* [26]. Also, a deletion of *FAM69B* in autism has been observed [27]. A network-based analysis of genes with CNV in autism identified involvement of synapse formation and function processes [28].

In the dbSNP database at NCBI, there is a total of 29 non-synonymous single nucleotide polymorphisms (SNPs), stop codon gain SNPs and frameshift SNPs in the five human FAM69 genes. Some of them may cause loss of the protein active site, may affect protein active site stability and conformation, protein localisation, calcium ion binding, or ATP-binding. However, with two exceptions (below), these SNPs are not linked to diseases and the allele frequencies are not known. Two polymorphisms, causing the S24P and K128R variation in CXORF36, are linked to Fragile X Syndrome [19, 20]. The first of these SNPs may affect protein localisation while the effect of the second, if any, cannot be predicted in the context of kinase-like structure model.

Discussion

We present the discovery of a novel 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]; [32].

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 conserved kinase function. For very distant homologues, the sequence alignment details are known to be less reliable than the overall detection of homology [35]. Thus, some of our definitions of FAM69 active site motifs or secondary structure element assignments may be erroneous. Also, it is not straightforward to predict the substrate.

Neurological disorders related to FAM69 genes may have a common denominator, malfunction of the secretory pathway in neurons [36]. Vesicular trafficking is critical in neuron development [37]. The calcium binding domain within the FAM69s agrees with the role of this ion in regulation of Golgi function [38]. FAM69s may be the yet unidentified kinases regulating of ER-to-Golgi vesicle transport [39].

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 will only come from experiments.

Methods

For remote homology identification, PSI-BLAST searches used the standard parameters on nr database at NCBI as of 05.2011. HHsenser [40] was used for surveying homologues, querying the nr database with standard parameters. For domain assignments, HMMER3 on the Pfam database as of 05.2011 was used.

For survey of similarities within the kinase-like clan, the CLANS algorithm [41] 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 (K. Pawlowski, in preparation), d) seeds from the Pfam families: Alpha_kinase (PF02816), PI3_PI4_kinase (PF00454), Act-Frag_cataly kinase, PF09192, PPDK_N (PF01326), PIP5K (PF01504). For the latter, 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 graph, similarity relations with significance of P-value below 0.1 were considered.

Transmembrane region predictions were achieved by the TMHMM and Phobius servers. The Jpred and PsiPred servers were used to predict the secondary structures. The multiple alignment was built using the MUSCLE program.

For three-dimensional structure prediction, two methods were used, namely FFAS03 [42], that uses sequence profile-to-profile comparison and HHpred [43] that employs HMM-to-HMM comparison. Both methods queried the Pfam, PDB and SCOP databases.
Sequence alignments between the kinase-like region of FAM69A and the PKA structure [PDB:1cdk] were produced by HHpred structure prediction method and manually adjusted to accommodate predicted secondary structures. 3D structure models were constructed automatically by the program MODELLER [44]. The ATP analogue and two Mn^{2+} ions from the 1cdk structure were replaced by ATP and two Mg^{2+} ions. The MetaMQAP server [45] was used to estimate the correctness of the 3D models using a number of model quality assessment methods in a meta-analysis.

List of abbreviations used

FAM69: Family with sequence similarity 69

PKA: Protein kinase A

PKL: Protein kinase-like

FXS: fragile X syndrome

XLMR: X-linked mental retardation

Authors' contributions

KP discovered the FAM69 similarity to kinases. KP, MD, AL carried out the bioinformatics analyses. KP conceived the study, designed and coordinated it. KP drafted the manuscript.

Acknowledgements

We thank Dr Marcin Grynberg for critical reading of the manuscript and helpful comments.

References

1. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S: **The protein kinase complement of the human genome**. *Science* 2002, **298**(5600):1912-1934.
2. Kannan N, Taylor SS, Zhai Y, Venter JC, Manning G: **Structural and functional diversity of the microbial kinome**. *PLoS Biol* 2007, **5**(3):e17.
3. Cheek S, Ginalski K, Zhang H, Grishin NV: **A comprehensive update of the sequence and structure classification of kinases**. *BMC Struct Biol* 2005, **5**:6.
4. Taylor SS, Radzio-Andzelm E: **Three protein kinase structures define a common motif**. *Structure* 1994, **2**(5):345-355.
5. Manning G, Plowman GD, Hunter T, Sudarsanam S: **Evolution of protein kinase signaling from yeast to man**. *Trends Biochem Sci* 2002, **27**(10):514-520.
6. Eglen R, Reisine T: **Drug discovery and the human kinome: Recent trends**. *Pharmacol Ther* 2011.
7. Manning BD: **Challenges and opportunities in defining the essential cancer kinome**. *Sci Signal* 2009, **2**(63):pe15.

8. Fedorov O, Muller S, Knapp S: **The (un)targeted cancer kinome**. *Nat Chem Biol* 2010, **6**(3):166-169.

9. Edwards AM, Isserlin R, Bader GD, Frye SV, Willson TM, Yu FH: **Too many roads not taken**. *Nature* 2011, **470**(7333):163-165.

10. Samir AA, Ropolo A, Grasso D, Tomasini R, Dagorn JC, Dusetti N, Iovanna JL, Vaccaro MI: **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* 2000, **4**(3):188-193.

11. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic Acids Res* 1997, **25**(17):3389-3402.

12. Aziz A, Harrop SP, Bishop NE: **Characterization of the deleted in autism 1 protein family: implications for studying cognitive disorders**. *PLoS One* 2011, **6**(1):e14547.

13. Tennant-Eyles AJ, Moffitt H, Whitehouse CA, Roberts RG: **Characterisation of the FAM69 family of cysteine-rich endoplasmic reticulum proteins**. *Biochem Biophys Res Commun* 2011, **406**(3):471-477.

14. Hanks SK, Quinn AM, Hunter T: **The protein kinase family: conserved features and deduced phylogeny of the catalytic domains**. *Science* 1988, **241**(4861):42-52.

15. Grabarek Z: **Structural basis for diversity of the EF-hand calcium-binding proteins**. *J Mol Biol* 2006, **359**(3):509-525.

16. Takatalo M, Jarvinen E, Laitinen S, Thesleff I, Ronnholm R: **Expression of the novel Golgi protein GoPro49 is developmentally regulated during mesenchymal differentiation**. *Dev Dyn* 2008, **237**(8):2243-2255.

17. Takatalo MS, Tummers M, Thesleff I, Ronnholm R: **Novel Golgi protein, GoPro49, is a specific dental follicle marker**. *J Dent Res* 2009, **88**(6):534-538.

18. Morrow EM, Yoo SY, Flavell SW, Kim TK, Lin Y, Hill RS, Mukaddes NM, Balkhy S, Gascon G, Hashmi A *et al*: **Identifying autism loci and genes by tracing recent shared ancestry**. *Science* 2008, **321**(5886):218-223.

19. Jensen LR, Lenzner S, Moser B, Freude K, Tzschach A, Wei C, Fryns JP, Chelly J, Turner G, Moraine C *et al*: **X-linked mental retardation: a comprehensive molecular screen of 47 candidate genes from a 7.4 Mb interval in Xp11**. *Eur J Hum Genet* 2007, **15**(1):68-75.

20. Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, O'Meara S, Latimer C, Dicks E, Menzies A *et al*: **A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation**. *Nat Genet* 2009, **41**(5):535-543.

21. Hagerman R, Hoem G, Hagerman P: **Fragile X and autism: Intertwined at the molecular level leading to targeted treatments**. *Mol Autism* 2010, **1**(1):12.

22. Thiselton DL, McDowall J, Brandau O, Ramser J, d'Esposito F, Bhattacharya SS, Ross MT, Hardcastle AJ, Meindl A: **An integrated, functionally annotated gene map of the DXS8026-ELK1 interval on human Xp11.3-Xp11.23: potential hotspot for neurogenetic disorders**. *Genomics* 2002, **79**(4):560-572.

23. Good CD, Lawrence K, Thomas NS, Price CJ, Ashburner J, Friston KJ, Frackowiak RS, Oreland L, Skuse DH: **Dosage-sensitive X-linked locus influences the development of amygdala and orbitofrontal cortex, and fear recognition in humans**. *Brain* 2003, **126**(Pt 11):2431-2446.

24. Wang KS, Liu XF, Aragam N: **A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder**. *Schizophr Res* 2010, **124**(1-3):192-199.

25. Alcina A, Fernandez O, Gonzalez JR, Catala-Rabasa A, Fedetz M, Ndagire D, Leyva L, Guerrero M, Arnal C, Delgado C *et al*: **Tag-SNP analysis of the GFI1-EVI5-RPL5-FAM69 risk locus for multiple sclerosis**. *Eur J Hum Genet* 2010, **18**(7):827-831.

26. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS *et al*: **Functional impact of global rare copy number variation in autism spectrum disorders**. *Nature* 2010, **466**(7304):368-372.

27. Sanders JA, al. e: **Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism**. *Neuron* 2011, **70**:863-885.

28. Gilman SR, Iossifov I, Levy D, Ronemus M, Wigler M, Vitkup D: **Rare de novo variants associated with autism implicate a large functional network of genes involved in formation and function of synapses**. *Neuron* 2011, **70**(5):898-907.

35. Jaroszewski L, Rychlewski L, Godzik A: **Improving the quality of twilight-zone alignments**. *Protein Sci* 2000, **9**(8):1487-1496.

36. Aziz A, Harrop SP, Bishop NE: **DIA1R is an X-linked gene related to Deleted In Autism-1**. *PLoS One* 2011, **6**(1):e14534.

37. Giannandrea M, Bianchi V, Mignogna ML, Sirri A, Carrabino S, D'Elia E, Vecellio M, Russo S, Cogliati F, Larizza L *et al*: **Mutations in the small GTPase gene RAB39B are responsible for X-linked mental retardation associated with autism, epilepsy, and macrocephaly**. *Am J Hum Genet* 2010, **86**(2):185-195.

38. Dolman NJ, Tepikin AV: **Calcium gradients and the Golgi**. *Cell Calcium* 2006, **40**(5-6):505-512.

39. Nakagawa H, Miyazaki S, Abe T, Umadome H, Tanaka K, Nishimura K, Komori M, Matsuo S: **H89 sensitive kinase regulates the translocation of Sar1 onto the ER membrane through phosphorylation of ER-coupled beta-tubulin**. *Int J Biochem Cell Biol* 2011, **43**(3):423-430.

40. Soding J, Remmert M, Biegert A, Lupas AN: **HHsenser: exhaustive transitive profile search using HMM-HMM comparison**. *Nucleic Acids Res* 2006, **34**(Web Server issue):W374-378.

41. Frickey T, Lupas A: **CLANS: a Java application for visualizing protein families based on pairwise similarity**. *Bioinformatics* 2004, **20**(18):3702-3704.

42. Rychlewski L, Jaroszewski L, Li W, Godzik A: **Comparison of sequence profiles. Strategies for structural predictions using sequence information**. *Protein Sci* 2000, **9**(2):232-241.

43. Soding J, Biegert A, Lupas AN: **The HHpred interactive server for protein homology detection and structure prediction**. *Nucleic Acids Res* 2005, **33**(Web Server issue):W244-248.

44. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints**. *J Mol Biol* 1993, **234**(3):779-815.

45. Pawlowski M, Gajda MJ, Matlak R, Bujnicki JM: **MetaMQAP: a meta-server for the quality assessment of protein models**. *BMC Bioinformatics* 2008, **9**:403.

46. Knighton DR, Zheng JH, Ten Eyck LF, Ashford VA, Xuong NH, Taylor SS, Sowadski JM: **Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase**. *Science* 1991, **253**(5018):407-414.

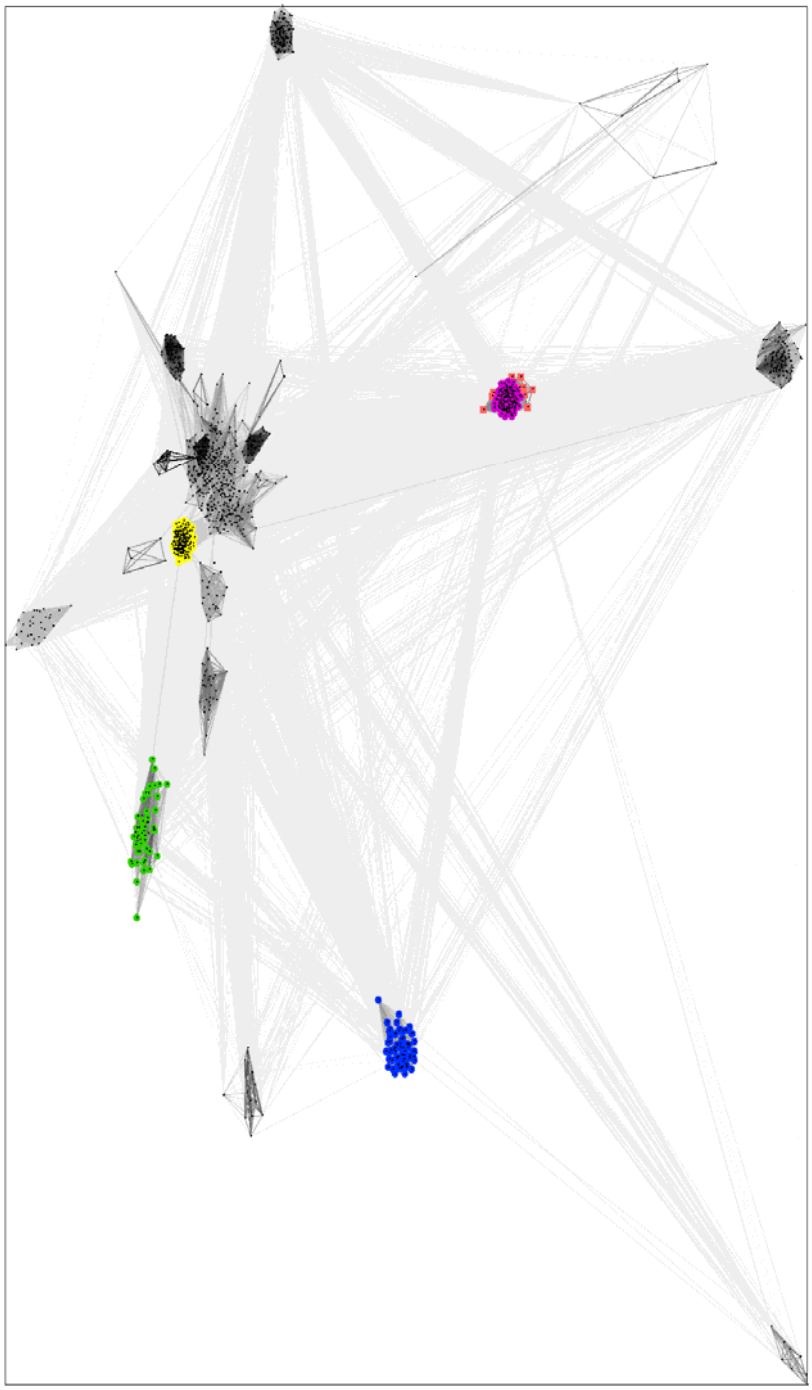
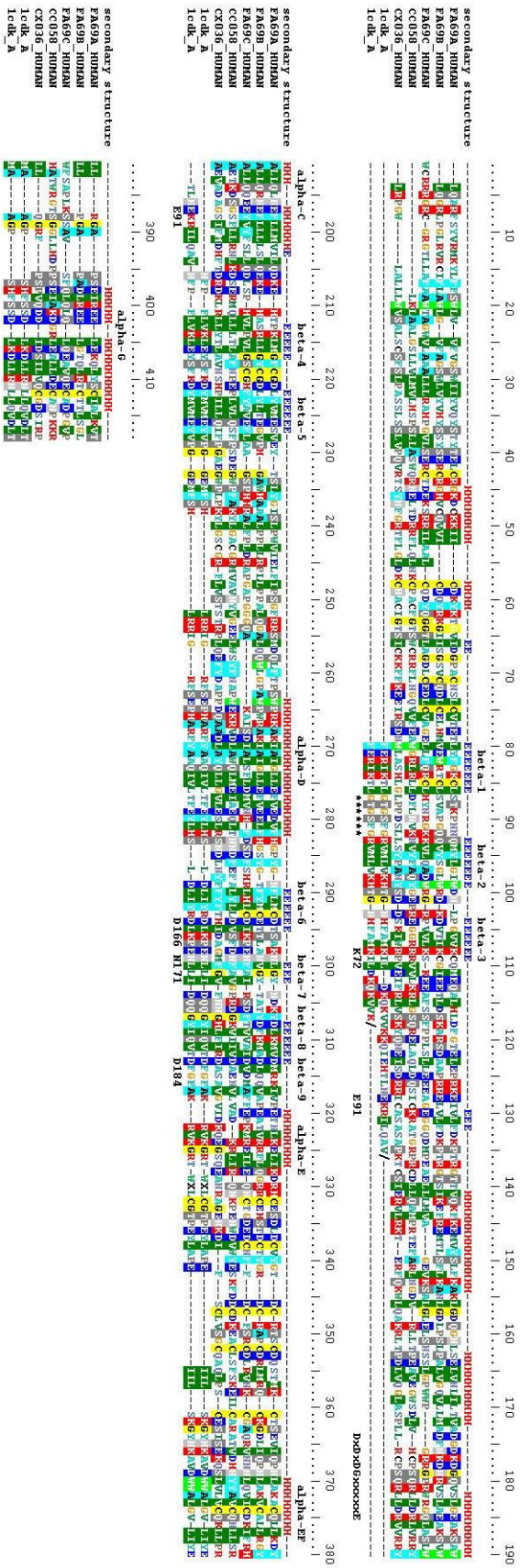


Fig. 1

CLANS graph visualizing PSI-BLAST-detected significant (dark grey) and sub-significant (light grey) similarities among protein kinase-like proteins. Green: FAM69 family, Yellow and brown: pkinase and pkinase_Tyr families, Blue: Alpha kinases, magenta: SELO, Black: other kinase families.

Figure 2.

Multiple sequence alignment of selected FAM69s covering the region 17-399 of FAM69A. Secondary structure prediction for human FAM69A protein shown. Secondary structure elements named as in PKA [46]. Locations of predicted key catalytic residues shown, in standard PKA numbering (e.g. D166), as well as the EF-hand motif and the ATP-binding loop (asterisks). PKA sequence is aligned using HHpred alignment to FAM69A.



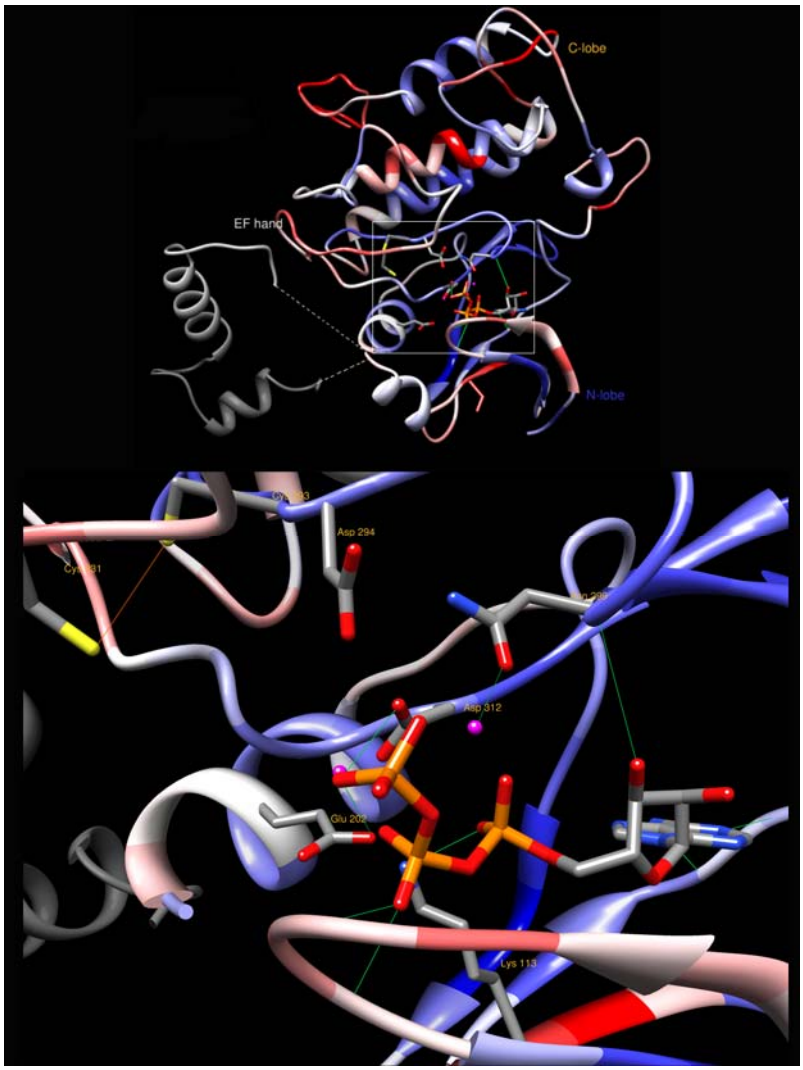


Figure 3.

Structure model of the kinase domain of human FAM69A. *Top*. Model coloured by MetaMQAP model quality score (blue: good quality, red: poor quality). On left, the EF-hand motif is shown in an approximate location. *Bottom*. As in top panel, close-up view of predicted active site with ATP molecule and two Mg^{2+} ions. Side chains of key active site residues shown: D294 (PKA numbering: 166), N299 (171), D312 (184), also the two cysteines near the active site: C293 and C331.