Analysis of extended kinome of zebrafish with immensely over represented PIM kinase subfamily

R. Rakshambikai, N. Srinivasan and Rupali A. Gadkari

Abstract:
In the recent times, zebrafish has gained lot of popularity as a model organism to study human cancers experimentally. Despite the phylogenetic distance between the two organisms zebrafish develops almost all types of human tumors when induced and the availability of transparent transgenes make the study of complex processes namely cancer metastasis relatively easier. However, there exist lacunae in understanding the mechanistic details of the tumour development due to scanty information available about the cellular components participating in various metabolic processes. Extensive analysis of the functional elements at the proteome level will add valuable information to fill the existing gap.
Present study was aimed at the function annotation of zebrafish proteome and comparative analysis of the functional elements with those found in humans. Towards this, the draft proteome sequence of zebrafish (ZV9) was analysed using remote homology detection tool HMMER3 using protein family (Pfam) profile sequences for comparison. The results revealed "substantial expansion" of the protein kinase family in zebrafish compared to humans, which constitutes over 3% of the entire proteome with 1200 members. These protein kinase like sequences were further mapped to individual kinase subfamilies on the basis of similarity to the kinase catalytic domain profiles. Motifs that are a pre-requisite for functional kinases were then identified in the shortlisted sequences to validate the hits obtained to be functionally active and these proteins were further compared to their human counterparts, as summarized in the Table 1. It is evident from the table that CAMK subfamily is hugely expanded in zebrafish while others look largely comparable.

Kinase	TK	TKL	AGC	CAMK	CK1	CMGC	STE	Other	Unclassified	Total
Zebrafish	165	79	114	465	20	115	78	110	54	1200
%	13.75	6.58	9.5	38.75	1.667	9.58	6.5	9.16	4.5	
Human										
%	20.34	7.2	11.72	16.72	1.8	19.82	8.9	9.0	4.27	

Over-representation of CAMK subfamily in zebrafish was attributed to the "massive" increase in the number of PIM group of kinases that play crucial roles in cell cycle regulation and growth, while most of the others groups of CAMK subfamily are marginally under represented.
Identities of the PIM kinases were further ascertained by the use of a two-way BLAST search using the SWISS-PROT database containing well annotated PIM kinases. The similarity between the well annotated sequences and the zebrafish PIM kinases were analysed by means of a dendrogram (Figure 1) which showed that only three of the PIM kinases cluster along with the well annotated PIM kinases (marked in red) indicating that probably only three of the PIM kinases function as the canonical PIM kinases (highlighted in green in the Figure). PIM kinases in other organisms are known to be constitutively active and there are residues that play crucial roles in maintaining the protein in functionally active conformation. To investigate the evolutionary conservation of these crucial residues, the protein sequences were then aligned using MAFFT algorithm and explored for key residues namely two proline residues at the hinge region along with Lys67. Interestingly, although the protein sequences seemed to have diverged, the residues important for maintaining active

conformation are highly conserved (scores of greater than 0.7). The substrate binding residues however, do not show high conservation indicating alternate substrate binding modes for the zebrafish kinases. Most of the interacting partners in the signalling by PIM kinases were identified by a stringent BLAST search and were found to be marginally overrepresented, equilibrating the increase in PIM kinases to some extent.

Activity of constitutively active PIM kinase is regulated largely at the level of translation of its mRNA as well as by maintaining short half-life of the protein. PIM kinase mRNA possesses a sequence motif "AUUUA" in its 3' untranslated region (3'UTR) which regulates its translation. Zebrafish PIM kinase RNA sequences were analysed subsequently. However, sequences of 3'UTRs could be obtained for only 20% of the PIM kinases, indicating possibility of other modes of regulation being active in the remaining proteins. The UTR sequences of 9 out of 36 sequences showed presence of above mentioned regulatory sequence motif out of which three sequences showed about 3-6 of these regulatory elements.

Interestingly these are the same three PIM kinase sequences of zebrafish that clustered closely with human PIM kinases (shown in Red in the Figure 1) as well as showed complete conservation of the substrate binding and other functional residues. This reiterates the possibility that this small group of proteins may be functioning as canonical PIM kinases. Roles played by other "PIM like kinases" need to be investigated further.

PIM kinases are known to be "weak oncogenic" proteins and are currently being targeted for the therapy of various human cancers. The present analysis of zebrafish PIM kinases has emerged various questions that need to be addressed in the field of cancer study and the "model organism" status of zebrafish needs to be substantiated. It would be extremely interesting and essential to understand if all the PIM kinases in zebrafish are activated when a tumor is induced.

Figure 1: Dendrogram with clustering of PIM kinases in zebrafish along with well annotated PIM kinases from other organisms. Colored in red are the well annotated PIM sequences from various organisms from SWISSPROT; shown in green are the sequences from zebrafish that cluster close to the classical PIM kinases and colored black are the zebrafish PIM kinases which deviate substantially from the classical PIM kinases. Part of the dendrogram shown in the right top of the figure corresponds to the zoomed in view of the part of the tree shown in the left highlighted by a square box.