

Delving into Vertebrate Serpins for Understanding their Evolution

Abhishek Kumar

Cellular Genetics, Faculty of Technology, University of Bielefeld, D-33501, Bielefeld, Germany

Email: abhishek.abhishekkumar@gmail.com

Abstract

The superfamily of serine proteinase inhibitors (serpins) is involved in an array of fundamental biological processes such as blood coagulation, cell differentiation, cell migration, complement activation, embryo implantation, fibrinolysis, angiogenesis, and inflammation, and tumor suppression. Vertebrate serpins can be conveniently classified into six sub-groups, based on three independent biological features - genomic organization, diagnostic amino acid sites and rare indels. The present vertebrate serpins are derived from an original serpin most probably by intron insertion and we are trying to reconstruct the phylogeny of vertebrate serpin and looking for the reconstruction of original vertebrate gene(s). We started with fish genomes and characterized fish serpins and assigned orthology with respect to human serpins. Most fish serpins are characterised as stereotype vertebrate serpins with some interesting exceptions which suggest that either there are some fish-specific serpins or some fish serpins do not have human orthologs.

Introduction

Serpins stand for **serine protease inhibitors** and there are around 500 serpins which from -- bacteria, viruses, plants and animals, which constitute a superfamily of proteins having low sequence similarity, but which are structurally conserved and often similar function. Serpins are generally 350-500 amino acids long, consisting of three beta-sheets and eight to nine alpha-helices, but the hallmark of serpins is 20 amino acid long C-terminal motif – reactive centre loop (RCL) where proteases bind and cleave and in turn get inactivated (1-2). Several aspects of the serpins suggest that they might provide an excellent vehicle for studying important questions about protein evolution, structure, and function

- **Extensive functional radiation among**

serpin proteins. Serpins regulate numerous separate intracellular and extracellular processes, including blood coagulation, cell differentiation, cell migration, complement activation, embryo implantation fibrinolysis, angiogenesis, inflammation, tumor suppression and other functions (3). While most serpins appear to act as protease inhibitors (4), some have lost this inhibitory role and function instead in blood pressure regulation (angiotensinogen) or hormone binding (corticosteroid-binding globulin).

- **There is considerable gene clustering.** For example, the genes coding for mammalian 1-antitrypsin like serpins and ov-serpins are clustered and organized into two clusters. The genes in these clusters lie in proximity on the same chromosome, suggesting that they might have arisen by tandem duplications from a common precursor (5-6). However, in spite of their close physical proximity, these proteins have disparate functions, raising interesting questions about the evolution of gene function and expression.
- **Serpin genes exhibit a variety of distinct exon-intron patterns.** The exon-intron structure of genes may contain important phylogenetic signals, and several authors have suggested the feasibility of evolutionary classifications based on comparative intron positioning (7-8). The serpin genes do exhibit differential exon-intron organization (9-12).
- **Serpins exhibit dynamic structural properties.** Serpins exist in active, cleaved, and latent forms (1).

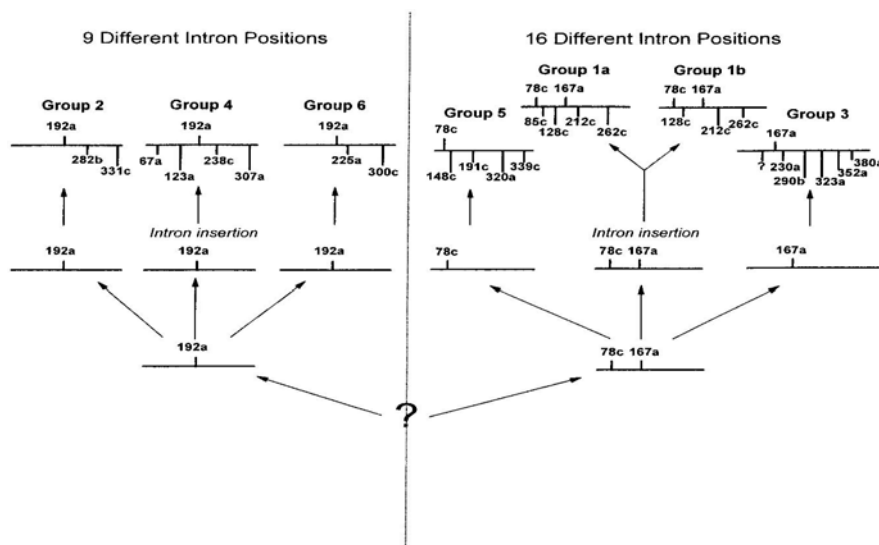


Fig 1. Phylogeny of vertebrate serpins based on gene structure, diagnostic amino acid sites, and indels suggest that present day vertebrate serpins have evolved from a single primordial gene by massive intron insertions (adopted from reference 12).

Thus, these properties suggest that serpins constitute an important group for comparative evolutionary analyses of protein structure and function. However, such studies require a well-documented phylogeny upon which to base speculations about structural and functional variation. The human genome encodes for at least 35 members that evolutionarily segregate into nine (A-I) distinct clades (13). Based on exon-intron organization, family-specific diagnostic amino acid sites, and rare indels, the vertebrate serpin evolution have been deduced (12), which suggested that present day vertebrate serpins can be classified into six groups which have arise from a single primordial gene by massive intron insertions before or during early vertebrate radiation as shown in figure 1.

In our work, we are trying to reconstruct the vertebrate serpin phylogeny and to find the answer to the following questions: (a) How and when did the supposed original vertebrate serpin gene(s) arise, (b) How and when did the six groups separate , (c) What is the mechanism of intron insertion in serpin genes? To answer of these questions, we started working on fish genomes like *Fugu*, zebrafish, and *Tetraodon* and we will move towards evolutionary important organisms, whose genomes are available, like chicken and *Xenopus*. Here, we are representing some of our results of studies on fish serpins, strategies in orthology assignments, and we seek your suggestions and ideas to carry out this project further.

Materials and Methods

Databases

We used the following fish genome databases: *Fugu* genome, *Danio* genome and *Tetraodon* genome. For differential comparison and annotation based data, we used NCBI (18), ENSEMBL (19) and SWISSPROT EXPASY (20) databases.

Searching Fish Serpins

Initially we looked at the genome databases of *Fugu*, *Danio* and *Tetraodon* for annotation of genomic sequences carrying serpin genes, but the quality of annotation was not good enough to have confidence of having all serpins from each fish genome. Therefore, we use the Genlight platform (21) for comparison and finding out all possible serpin genes using BLAST (22). Secondly, we scanned for three typical serpins motifs present in serpin sequences.

Gene Prediction

Gene structure predictions are mostly made through combination of GENSCAN (23-24), GenomeScan (23-24), and GeneWise (25) using genomic sequences.

Mapping intron positions

Intron positions of each serpin are mapped with respect to conserved regions of alpha-1-antitrypsin.

Orthology Assignments

Orthology is assigned with different independent parameters - percentage identity at protein sequence level, group-specific exon-intron structures, presence of gene specific signature motifs, by presence or absence of signal peptide, used group-specific diagnostic sites, reactive centre loop characteristics, Genlight's orthology finding tool, which uses bidirectional BLAST and synteny between human and fish genomes which gives idea of gene order on chromosome as serpins are found in a specific order in mammalian genomes.

For that serpin of interest are aligned with alpha-1-antitrypsin using ClustalW (26) and the introns in genomic sequences positioned with help of GeneWise (25) onto this aligned protein sequences.

Results and Discussions

We found 17 serpins in the *Fugu* genome which are also present in *Tetraodon* and zebrafish. Subsequently we assigned orthology to each fish serpin with human serpins. Most of time we found these have orthologs in human genome as listed in table 1. Fish serpins tend to classify in same fashion into 6 groups of serpins as mammalians, the exon-intron structure is similar as of higher vertebrates found from the same strategies used in reference 12.

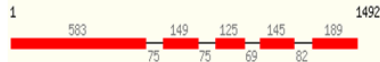
Table 1. List of *Fugu* serpins, with gene identifiers, scaffold Id where corresponding genes are found and possible orthology with human genes. Bold face indicates that the confidence in assigning orthology is high. Brackets organism name indicate ortholog found in that organism but not human ortholog.

S. No.	Fugu Serpents	Gene Id (Fugu v3 genome database)	Scaffold Id	Clone name (UK Fugu Database)	Human Ortholog
1	Fu-Spn-1	FRUP00000156735	scaffold_2913	FM:M002913	Heat Shock Protein 47
2	Fu-Spn-2	FRUP00000155065	scaffold_757	FM:M000757	Alpha-1-antitrypsin (<i>Xenopus</i>)
3	Fu-Spn-3	FRUP00000161527	scaffold_2007	FM:M002007	Neuroserpin precursor
4	Fu-Spn-4	FRUP00000132180	scaffold_188	FM:M000188	Z-dependent protease inhibitor
5	Fu-Spn-5	FRUP00000137160	scaffold_417	FM:M000417	Glia derived nexin precursor
6	Fu-Spn-6	FRUP00000149263	scaffold_385	FM:M000385	Heparin cofactor II
7	Fu-Spn-7	FRUP00000160285	scaffold_6239	FM:M006239	Alpha1-Antitrypsin
8	Fu-Spn-11	FRUP00000140727	scaffold_508	FM:M000508	Angiotensinogen
9	Fu-Spn-12	FRUP00000141273	scaffold_1026	FM:M001026	Pigment epithelium-derived factor
10	Fu-Spn-14	FRUP00000162952	scaffold_154	FM:M000154	Alpha-2-antiplasmin precursor
11	Fu-Spn-17	FRUP00000155064	scaffold_757	FM:M000757	Antitrypsin
12	Fu-Spn-18	FRUP00000146289	scaffold_641	FM:M000641	Antitrypsin
13	Fu-Spn-35	FRUP00000131353	scaffold_110	FM:M000110	Human LE1
14	Fu-Spn-36	FRUP00000163136	scaffold_2405	FM:M002405	Human LE1
15	Fu-Spn-37	FRUP00000138778	scaffold_5139	FM:M005139	Human LE1
16	Fu-Spn-38	FRUP00000165249	scaffold_1226	FM:M001226	Antithrombin III
17	Fu-Spn-39	FRUP00000133449	scaffold_3139	FM:M003139	C1-Inhibitor

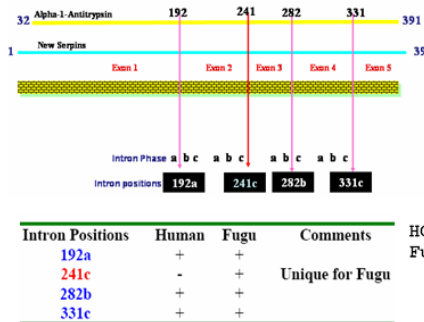
(a) Percent sequence identity = 53

(b) Signature motifs: Acidic repeats and conserved helix-D.

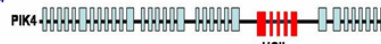
(c) Exon-intron Structure.



(d) Mapping of intron positions for HCII.



(e) HCII lies in a big intron of PIK4.



(f) Similar RCL.

```

P1P1'
  ||
G T Q A T T V T T V G F M P L S T Q V 19
G T E A A A L T Q V G F M P L S S Q I 19
* * : * : : * * * * * * * : *

```

Normally we achieved a good confidence level by assigning orthology with help of several independent parameters as shown in figure 2 - the case of Fu-Spn-6, which has human heparin cofactor II (HCII) as an ortholog, as indicated by 53% sequence identity, presence of acidic repeats and conserved helix-D which are functionally very important, exon-intron structure similarities. In human the HCII gene lies in a bigger intron of a gene called phosphatidylinositol 4-kinase (PIK4), same is in case of fishes as well, we used this as marker for HCII orthology. Finally of Fu-Spn-6 has very similar RCL as of HCII of human. These parameters indeed suggest that human HCII is ortholog of Fu-Spn-6. In some cases we have found human orthologs very easily as we had many parameters about genes. But in some cases especially in group 2 members, it was tough to indicate orthologs. In some cases there is no human ortholog like for Fu-Spn-2 which has alpha-1-antitrypsin like gene from *Xenopus* as ortholog which has extra His-rich domain which helps in estrogens binding. We found that fishes have serpins which are classified into 6 groups as in higher vertebrates. There are just 17 serpin genes in fishes, while 35 in human, 59 in mouse and 42 in rat, suggest that serpin genes don't follow whole genome duplication in fishes during early vertebrate evolution (28).

There are there serpin genes in fishes which are orthologous to Leukocyte elastase inhibitor of group 1, which is highly expanded by tandem gene duplication on the same chromosomes in human and still further expanded in rat and mouse (29). This is a very surprising result that fishes have just one of these ov-serpins while in mammals there are wide array of ov-serpins which are functionally highly diversified. Group 2 serpins in fishes are again just seven while there are more in mammals that suggest serpins are differentially expanded into different organisms supposedly because of organism based requirements.

Conclusions

The fish serpins tend to group in a similar fashion as higher vertebrates, since they have same exon-intron structures and conserved intron positions with some interesting exceptions which might be interesting clues to serpins evolution which is under investigation. Assigning orthology to serpin gene is not easy and it requires a big amount of gene specific information. By now we have assigned orthology to fish serpins and analysed these genes and it seems our proposed vertebrate serpin gene is evolved before fish and higher vertebrate separation around 450 My ago (where fish and mammals separated) and serpin genes do not follow whole genome duplication in fishes during early vertebrate evolution.

Fig 2. Assigning Orthology for Fu-Spn-6 based of following independent parameters: (a) It has 53% sequence identity with human heparin cofactor II (HCII). (b) Also carrying two signature motif of heparin cofactor II – acidic repeats and conserved helix-D. (c) Exon-intron structure is characteristic for group 2 members. (d) Intron phasing is denominated as a, b and c for first, second and third nucleotide from where introns are spliced out. Table shows the intron positions in Fu-Spn-6 which is similar to human HCII except for 241c. (e) The human HCII gene lies within bigger intron of another gene called PIK4 and this situation is found in fishes as well. (f) Fu-Spn-6 has very similar RCL motif as human HCII. Therefore, Fu-Spn-6 is indeed the ortholog of human HCII.

References

- Irving, J. A., Pike, R. N., Lesk, A. M., & Whisstock, J. C. (2000). Phylogeny of the serpin superfamily: implications of patterns of amino acid conservation for structure and function. *Genome Research*, **10** (12), 1845–1864.
- Gettins G.W.P, (2002). Serpin structure, Mechanism, and Function. *Chem Review* **102**, 4751-4803.
- Potempa, J., Korzus E., and Travis. J. (1994). The serpin superfamily of proteinase inhibitors: structure, function, and regulation. *J. Biol. Chem.* **269**:15957–15960
- Wilczynska M., M. Fa, P. I. Ohlsson, T. Ny, (1995). The inhibition mechanism of serpins. Evidence that the mobile reactive center loop is cleaved in the native protease-inhibitor complex *J. Biol. Chem* **270**:29652-29655
- Billingsley G. D., M. A. Walter, G. L. Hammond, D. W. Cox, (1993). Physical mapping of four serpin genes: alpha 1-antitrypsin, alpha 1-antichymotrypsin, corticosteroid-binding globulin, and protein C inhibitor, within a 280-kb region on chromosome I4q32.1 *Am. J. Hum. Genet* **52**:343-353
- Rollini P., R. E. Fournier, (1997). A 370-kb cosmid contig of the serpin gene cluster on human chromosome 14q32. 1: molecular linkage of the genes encoding alpha 1-antichymotrypsin, protein C inhibitor, kallistatin, alpha 1-antitrypsin, and corticosteroid-binding globulin. *Genomics*, **46**:409-415
- Long M., S. J. de Souza, W. Gilbert, (1995). Evolution of exon-intron structure of eukaryotic genes *Curr. Opin. Genet. Dev* **5**:774-778
- Logsdon J. M. Jr., A. Stolfus, W. F. Doolittle. (1998) Molecular evolution: recent cases of spliceosomal intron gain. *Curr. Biol* **8**:560-563
- Bao, J.-J., R. N. Sifers, V. J. Kidd, F. D. Ledley, and S. L. C. Woo. (1987). Molecular evolution of serpins: homologous structure of the human α_1 -antichymotrypsin and α_1 -antitrypsin genes. *Biochemistry* **26**:7755–7759.
- Ragg H., G. Preibisch, (1988). Structure and expression of the gene coding for the human serpin hLS2. *J. Biol. Chem* **263**:12129-12134
- Remold-O'Donnell E., (1993).The ovalbumin family of serpin proteins *FEBS Lett* **315**:105-108
- Ragg H., T. Lokot, P. B. Kamp, Atchley, W.R., Dress A.(2001) Vertebrate serpins: construction of a conflict-free phylogeny by combining exon-intron and diagnostic site analyses *Mol. Biol. Evol* **18**:577-584
- Silverman, G. A., Bird, P. I., Carrell, R. W., Church, F. C., Coughlin, P. B., Gettins, P. G. W., Irving, J. A., Lomas, D. A., Luke, C. J., Moyer, R. W., Pemberton, P. A., Remold-O'Donnell, E.,Salvesen, G. S., Travis, J., and Whisstock, J. C. (2001). The serpins are an expanding superfamily of structurally similar but functionally diverse proteins. *Journal of Biological Chemistry*, **276**, 33293–33296.
- Fugu genome v3 database, <http://genome.jgi-psf.org/fugu6/fugu6.home.html>.
- Fugu genome sequencing project, <http://fugu.hgmp.mrc.ac.uk/>.
- Danio rerio* genome, http://www.sanger.ac.uk/Projects/D_rerio/.
- Tetraodon* genome, <http://www.genoscope.cns.fr/externe/tetranew/dat a/assembly/>.
- National center for biotechnological information, <http://www.ncbi.nlm.nih.gov/>.
- Ensemble, <http://www.ensembl.org/>.
- Swissprot, <http://us.expasy.org/sprot/>.
- Genlight, <http://piranha.techfak.uni-bielefeld.de/>
- Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403-10.
- Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78-94.
- Burge, C. B., and Karlin, S. (1998). Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* **8**, 346-354.
- Genewise, <http://www.ebi.ac.uk/Wise2/#>.
- Higgins D., Thompson J., Gibson T. Thompson J.D., Higgins D.G., Gibson T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
- Yoshida, K., Suzuki, Y. and Sinohara, H. (1999) Cloning and comparative sequence analysis of *Xenopus laevis* alpha1-antiproteinase. *J. Biochem. Mol. Biol. Biophys.* **3**, 59-63.
- Ohno, S. (1970). *Evolution by gene duplication*. Springer-Verlag, NY.
- Puente XS, Lopez-Otin C. A (2003). Genomic analysis of rat proteases: a comparative genomic approach. *Nat Rev Genet.* **4**, 544-58.