



## Prox1 Gene Characterization and Mutational Analysis

R. Lakshita

SRM Institute of Science and Technology, Chennai, Tamil Nadu, India

Phone: 09884477171, Email ID: lakshita07@gmail.com

Received: Aug 20<sup>th</sup> 2012, Accepted: Aug 26<sup>th</sup> 2012, Published: Sep 1<sup>st</sup> 2012

### Abstract:

Prox1 gene encodes for the homeo domain protein. It is the mammalian homologue of the gene prospero in *Drosophila*. The gene has a vital role in the development of brain structure and is crucial for learning and memory. In view of its importance the present work aims to use the Insilico tools and softwares to characterize this gene and the corresponding protein with special emphasis on the mutational sites. The gene and the protein sequences were retrieved from the NCBI data base for their further analysis. The gene characterization includes Promoter site identification, poly A site and Splice site detection, Motif analysis, prediction of the conserved sites. CpG Islands present in the gene sequence were predicted so as to find the most stable regions of the gene. Translation initiation sites were identified that can be help in predicting the frame of the gene product. This was followed by a focus on the mutation site prediction and analyzing its importance dictating the gene product. Possible Mutations in the prox1 gene are studied from SNPdb and Ensembl. The effects of these mutations on the translated product i.e, protein and its stability are analyzed using I-Mutant. SIFT was used to check the effect of these mutations on the normal function of protein. The effect of mutations on the onset of diseased condition was traced using PhdSNP. The ultimate outcome of the work is to establish the biologically important sites within the gene that can have a direct role in regulating the function of the gene product. The work can be a considerable support for the future research on the Prox 1 gene.

### Introduction:

The homeobox gene *prox1* is related to the *Drosophila prospero* gene, which mediates cell fate decisions of neuroblasts <sup>[1][2]</sup>. Prox1 has been found in many species such as *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Gallus gallus* and *Danio rerio*. *Prox1* is the master gene of lymph angiogenesis <sup>[3]</sup> but is also expressed in the developing central nervous system <sup>[4]</sup>, lens-secreting cone cells<sup>[5][6]</sup>, R7 photoreceptors, and midgut. Furthermore, the expression is detected in embryonic hepatoblasts and in adult hepatocytes <sup>[7][8]</sup>. Prox1 controls cell proliferation and differentiation through different transcription pathways and has both oncogenic and tumor-suppressive functions.

Human *Prox1* gene contains 736 amino acids. It contains two highly conserved domains one homeodomain and one prospero domain. *Prox1* is a key player in normal development of a brain structure crucial for learning and memory and remains active throughout life; nurturing the cells vital for making new memories <sup>[9]</sup>. *Prox1* is a transcription factor that functions like an on-off switch for genes. Researchers showed that by removing *Prox1* at different stages of mouse development, the dentate gyrus fails to develop properly. Investigators also demonstrated that *Prox1* remains important throughout mammalian life to ensure production of new granule cells, which are needed to form new memories. In this study, investigators determined that in the adult mammalian brain, *Prox1* is active during a particular stage of differentiation when neural stem cells change from cells of unlimited potential to more specialized granule cells. The scientists reported that in the dentate gyrus Prox1 is produced by intermediate progenitor cells and that the lack of Prox1 results in death of these

intermediate cells. Without these intermediate progenitor cells, new adult granule cells do not develop. Prox1 is necessary for the differentiation of granule cells and reveals a regulatory mechanism that links the production of the proper number of new granule cells to the maintenance of adult stem cells.

**Materials and Methods:**

The sequence (Human homeodomain protein (Prox 1) mRNA, complete cds) was retrieved from NCBI. Sequence (Gene and Protein) Analysis was performed for finding the PROX1 homology in different vertebrates. The organisms considered for the homology study were: Mus musculus, Rattus norvegicus, Macaca mulatta, Sus scrofa, Gallus gallus, Danio rerio and Carassius auratus. BLAST tool was used to check the degree of homology among the sequences based on pair wise sequence alignment. To analyze the evolutionary relationship among the selected sequences (both the protein and gene separately) Clustal W tool based on Multiple Sequence Alignment was used. Other tools of SDSC Biology work bench like Texshade and Boxshade were employed to analyze the patterns of conservations among the aligned sequences. The conservation analysis helps to understand the significant sites within the sequences that have been remained unchanged from several generations. Complete Gene annotation was performed using various Insilico tools.

Potential exonic regions in the gene sequence were detected using FEX. Tools like FSplice, PolyAH, CpG Island were used to study the Splice sites, Poly adenylation sites and regions rich in GC content respectively. Comet was used to identify the Cis Elements which are of regulatory function in the gene. The domain regions were identified in the protein using ProDom database. SNP database and Ensemble were used to predict the Mutations in the gene sequence. I-mutant2.0 was used to analyze the affect of these mutations on the stability of the protein. To identify whether the

mutation was deleterious SIFT was used. The influence of this mutation on the onset of disease is indicated by the tool PhdSNP.

**Results:**

The sequence, Human homeodomain protein (Prox 1) mRNA, complete cds was retrieved from NCBI.

BLAST RESULT: BLAST analysis to find the degree of homology among the selected sequences showed the following results:

- 100% identity to Human homeodomain protein (Prox 1) mRNA, complete cds
- 90% identity to Mus musculus prospero-related homeobox 1 (Prox1), mRNA
- 89% identity to Rattus norvegicus prospero homeobox 1 (Prox1), mRNA
- 99% identity to Macaca mulatta prospero homeobox 1 (PROX1), mRNA
- 96% identity to Sus scrofa prospero homeobox 1 (PROX1), mRNA
- 89% identity to Gallus gallus prospero homeobox 1 (PROX1), mRNA
- 77% identity to Danio rerio prospero-related homeobox gene 1a (prox1a), mRNA
- 76% identity to Carassius auratus prospero-related homeobox protein (prox1) mRNA, complete cds

It can be stated from the above results that the Prox1 gene has high degree of homology with the selected sequences.

**Identification of Sequence Conservation:**

Using TEXSHADE the conserved regions in the PROX1 coding protein were identified.

**Table 1: The conserved regions in human protein**

REGION	PATTERN
281-284	SFIL
319-322	EMAE
331-336	NKERDH
425-429	GNVQM
500-504	SGSFS
574-596	AMQEG.....YTRY
598-736	SSNMLK.....LLHE

The numeric regions in the above table indicate the location of the conserved site with respect to Human Prox1

**Table 2: The conserved regions in the human Prox1 gene**

REGION	PATTERN
551- 597	tcttg.....taac
601-606	ccagtg
701-703	cat
712-714	aag
1038-1040	tgg
1162-1167	gggcaat
1182-1189	gatggccc
1466-1472	gggccca
1739-1741	agg
1771-1780	cccaggcca
1887-1891	gcaga
2028-2032	cacca
2108-2114	gctcctt
2217-2225	gttcaccag
2562-2566	gtcta

**Exon Site Prediction:**

Two potential exonic regions were found using FEX.

- One of the exon is in the region 277-449
- Second exon is at 739-2817

**Poly A tail Prediction:**

There was one poly A site predicted in the sequence which is at 2611 position that has its role in providing stability to the mRNA CpG Islands were maximum concentrated in the region 248 to 466 which indicates that this is the most stable region in the gene that has high GC Content.

**Cis elements:**

The following Motifs were the Cis elements found in the Prox1 gene sequence: TATA, GATA, LSF, Myf and Ets. These elements have a vital role in gene expression and regulation.

Using TIS tool, possible translation initiation sites were identified at positions 739,607,907,859 and 856.

**Domain analysis:**

Five Domain regions were identified from positions using ProDom tool: 1-143, 144-428, 429-564, 545-606, and 608-733.

**Table 3: Mutations reported in SNPdb and ENSEMBL**

Nucleotide position	Mutation	Amino acid position	Amino acid variation
701	C/A	32	A/E
977	A/G	124	H/R
1162	G/C	186	G/S
1280	T/C	225	V/A
1445	G/A G/C	280	R/H R/P
1508	A/G	301	E/G
1613	A/C	336	H/P

The above mentioned sites can be considered to be important in the mutation study of the Prox1 gene.

**Table 4: I-MUTANT 2.0 RESULTS**

Position	Wild type	New	Stability	RI	pH	T
32	A	E	Decrease	8	7.0	25
124	H	R	Decrease	8	7.0	25
186	G	S	Decrease	8	7.0	25
225	V	A	Increase	4	7.0	25
280	R	H P	Decrease Decrease	8 5	7.0	25
301	E	G	Decrease	3	7.0	25
336	H	P	Increase	3	7.0	25

The above table shows that the substitution of V225 A and H336P are favorable as they increase the stability of the protein.

**Table 5: SIFT Results in identifying the deleterious mutations**

Position	Wild type	New	Effects	SIFT Value
32	A	E	Deleterious	0.01
124	H	R	Tolerated	0.2
186	G	S	Deleterious	0.05
225	V	A	Tolerated	1
280	R	H P	Tolerated Deleterious	0.12 0.02
301	E	G	Tolerated	0.08
336	H	P	Tolerated	0.2

The above results indicate that the mutations A32E, G186S and R280P are deleterious when compared to the other mutations reported in SNP Database.

**Table 6: PHD-SNP Results to identify the disease causing mutations**

Position	Wild type	New	Effects	RI
32	A	E	Disease	0

From the above table it can be inferred that mutations at the position 32 will lead to a diseased condition when compared to all the other mutations as shown by I mutant and SIFT tools.

According to I mutant the mutations at positions 32,124,186,280,301 decrease the stability of the protein. These mutations were further analyzed using SIFT which showed that the positions 32, 186 and 280 are deleterious. From among these deleterious mutations the site 32 was of more importance since the PhdSNP tool has shown the site to be more susceptible in causing the onset of a disease.

**Discussion and Conclusion:**

Prox1 gene of Homo sapiens was characterized. Gene annotation was done. Two potential exon regions were found which include 277-449 and 739-2817. Potential polyA site was predicted at

position 2611 which is the site responsible for maintaining the stability of the gene. CpG Island was detected in the region 248 to 466 which infers that this is the region that is most stable in the entire sequence. Several Cis elements like TATA, GATA, LSF, Myf and Ets were found using Comet. Possible translation initiation sites were identified at positions 739,607,907,859 and 856 of PROX1 gene. Five domain regions were identified at the positions 1-143, 144-428, 429-564, 545-606, 608-733 of protein. Mutations were studied using Ensembl and SNPdb. The mutations at positions 32,124,186,280,301 decrease the stability of the protein. The mutations which had deleterious effects were studied using SIFT. The mutations at 32,186 and 280 were found have deleterious effects.

The mutation of C to A at the position 701 in the conserved region of gene would result in coding of glutamate instead of alanine in the protein at the position 32<sup>nd</sup>. The deleterious mutations 32,186 and 280 were further analyzed using PhdSNP, the effect of the mutation at position 32 is leading to the diseased condition. This mutation may affect the structural and functional properties of the protein affecting the CNS of Homo sapiens as PROX1 gene plays a major role in development of CNS. The work has highlighted all the regions of structural, functional and evolutionary importance in the gene as well as protein sequence of Prox 1 which can be a future scope for research in this domain.

**Acknowledgement:**

I would like to thank my college SRM INSTITUTE OF SCIENCE AND TECHNOLOGY for providing me with this opportunity to enhance my skills and bring my knowledge into practicality.

**References:**

1. Doe CQ, Chu-LaGraff Q , Wright DM, Scott MP, The *prospero* gene specifies cell fates in the drosophila central nervous system. Cell, Volume 65, Issue 3, 451-464, 3 May 1991. doi:10.1016/0092-8674(91)90463-9
2. Chu-Lagraff, Q., Wright, D. M., McNeil, L. K., and Doe, C. Q. (1991) The *prospero* gene encodes a divergent homeodomain protein that controls neuronal identity in *Drosophila*. Development (Suppl.) 2: 79–85.
3. Hong, Y.-K., Harvey, N., Noh, Y.-H., Schacht, V., Hirakawa, S., Detmar, M. and Oliver, G. (2002), Prox1 is a master control gene in the program specifying lymphatic endothelial cell fate. Dev. Dyn., 225: 351–357. doi: 10.1002/dvdy.10163
4. Lavado, A. and Oliver, G. (2007), *Prox1* expression patterns in the developing and adult murine brain. Dev. Dyn., 236: 518–524. doi: 10.1002/dvdy.21024
5. Tomarev, S. I., Sundin, O., Banerjee-Basu, S., Duncan, M. K., Yang, J.-M. and Piatigorsky, J. (1996), Chicken homeobox gene prox 1 related to *Drosophila prospero* is expressed in the developing lens and retina. Dev. Dyn., 206: 354–367. doi: 10.1002/(SICI)1097-0177(199608)206:4<354::AID-AJA2>3.0.CO;2-H
6. Jeffrey T. Wigle, Kamal Chowdhury, Peter Gruss & Guillermo Oliver *Prox1* function is crucial for mouse lens-fibre elongation Nature Genetics 21, 318 - 322 (1999) doi:10.1038/6844
7. Maria Papoutsi, Jozsef Dudas, Jürgen Becker, Marco Tripodi and Lennart Opitz, et al. Gene regulation by homeobox transcription factor Prox1 in murine hepatoblasts Cell and Tissue Research, 2007, Volume 330, Number 2, Pages 209-220
8. J. Dudas, M. Papoutsi, M. Hecht, A. Elmaouhoub, B. Saile, B. Christ, S. I. Tomarev, C. S. von Kaisenberg, L. Schweigerer and G. Ramadori, et al. The homeobox transcription factor Prox1 is highly conserved in embryonic hepatoblasts and in adult and transformed hepatocytes, but is absent from bile duct epithelium. Anatomy and Embryology Volume 208, Number 5(2004), 359-366, DOI: 10.1007/s00429-004-0403-4
9. Lavado A, Lagutin OV, Chow LML, Baker SJ, Oliver G (2010) Prox1 Is Required for Granule Cell Maturation and Intermediate Progenitor Maintenance During Brain Neurogenesis. PLoS Biol 8(8): e1000460. doi:10.1371/journal.pbio.1000460

\*\*\*\*\*