International Journal of Advanced Research in Engineering and Applied Sciences

ISSN: 2278-6252

OMICS OF SERPIND1 FROM HOMO SAPIENS

Rizwana Begum S*

Manjula Bandari*

Swaroop Kumar RS*

Jagadeesh Kodavaty**

Abstract: *The product encoded by SERPIND1 gene is Heparin cofactor II (HCII), a serine proteinase inhibitor which rapidly inhibits thrombin in the presence of dermatan sulfate or heparin. Mutations in this gene are associated with heparin cofactor II deficiency. Heparin Cofactor II deficiency can lead to increased thrombin generation and a hypercoagulable state. The exploration of SERPIND1 gene with genomic and proteomics studies reveals that, the SERPIND1 is a pyramidine rich nucleotide sequence with 2237 base pairs with a molecular weight of 682937 Da (ssDNA). The primary structure of the coded peptide reveals that it is stable and acidic. Secondary structure defines that it mainly contains the random coils, alpha helix beta sheets and beta turn. The gene contains both left & right primers and several restriction sites for restriction enzymes. The protein has 20 phosphorylation sites (Serine, Threonine & Tyrosine), three net N-Glycosylation sites along with one O-Glycosylation (Threonine at position 104) site.*

Keywords*: SERPIND1, Heparin cofactor II (HCII), Homo sapiens, serine proteinase inhibitor, heparin cofactor II.*

*Department of Biotechnology, Jawaharlal Nehru Technological University Anantapur College of Engineering Pulivendula, Pulvendula, Andhra Pradesh, INDIA.

**Departments of Chemical Engineering, Indian Institute of Technology Madras, Chennai, INDIA.

INTRODUCTION

The official name of SERPIND1 gene is "serpin peptidase inhibitor, clade D (heparin cofactor), member1". The SERPIND1 gene is located on the long (q) arm of chromosome 22 at position 11.21 and code for the protein Heparin cofactor II (HCII). Defects in SERPIND1 are the cause of thrombophilia due to heparin cofactor 2 deficiency (THPH10), a hemostatic disorder characterized by a tendency to recurrent thrombosis **[1].** Serpins are a group of proteins with similar structures that were first identified as a set of proteins able to inhibit proteases. The acronym serpin was originally coined because many serpins inhibit chymotrypsin-like serine proteases (serine protease inhibitors) **[2,3,4]**. The first members of the serpin superfamily to be extensively studied were the human plasma proteins antithrombin and antitrypsin, which play key roles in controlling blood coagulation and inflammation, respectively. Initially, research focused upon their role in human disease: antithrombin deficiency results in thrombosis and antitrypsin deficiency causes emphysema **[5].** Over 1000 serpins have now been identified, these include 36 human proteins, as well as molecules in plants, fungi, bacteria, archaea and certain viruses **[6,7,8]**. Serpins are thus the largest and most diverse family of protease inhibitors **[9].** The product encoded by this gene is a serine proteinase inhibitor which rapidly inhibits thrombin in the presence of dermatan sulfate or heparin. Mutations in this gene are associated with heparin cofactor II deficiency. Heparin Cofactor II deficiency can lead to increased thrombin generation and a hypercoagulable state **[10]**.

METHODOLOGY

Various genomic & proteomic offline and online tools were used to explore the *SERPIND1* gene. The nucleic acid sequence of *SERPIND1 from Homo sapiens* was retrieved from NCBI (NM_000185.3). Genomic studies were done by Bioedit, ORF finder, Primer 3.0 and Genscan followed by proteomic studies to predict primary structure with Bioedit, Protparam, and secondary structure with SOPMA **[11]**. Post translational modifications are studied with NetOGlyc, NetNGlyc and NetPhos 2.0 tools.

RESULTS AND DISCUSSION

GENOMICS

Bioedit

DNA molecule: gi|73858565|ref|NM_000185.3| Homo sapiens serpin peptidase inhibitor, clade D (heparin cofactor), member 1 (SERPIND1), mRNA, Length = 2237 base pairs

Molecular Weight = 682937.00 Da, (ssDNA) & 1359559.00 Da, (dsDNA)

G+C content = 49.08% A+T content = 50.92%

Nucleotide	Number	Mol%
A	27	28.03
C	592	26.46
G	506	22.62
T	512	22.89

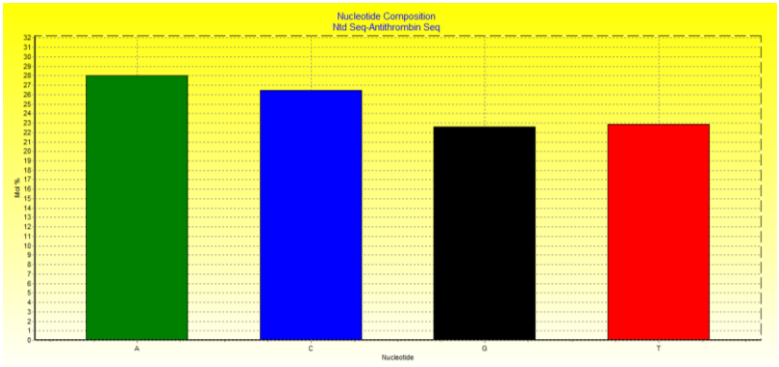

Figure 1 : Nucleotide composition of *SERPIND1 gene*

Nucleotide composition of SERPIND1 gene from Homo sapiens has been predicted by Bioedit & the results reveals that the sequence is pyramidine rich with molecular weight of more than 682937 Da (ssDNA) containing 2237 base pairs. The GC & AT content was found to be **49.08%** and **50.92%** respectively. The gene has a total of 22 restriction sites for the restriction enzymes (Prediction in view of Bangalore Genei)

ORF Finder

Frame	from	to	Length
+2	68	1567	1500
-2	1349	1774	426
-3	946	1221	276
-1	1980	2213	234
-3	1273	1503	231
-3	478	660	183
-3	796	945	150
-3	1	132	132
+1	2113	2235	123
-2	566	688	123
+1	1207	1323	117
-2	1835	1939	105

Open Reading Frame Finder predicts the presence of the possible open reading frames of the given sequence.

It was been identified that the SERPIND1 codes for 12 encoded proteins present in both the + and – strands.The largest ORF was identified in the 2nd frame of the direct strand from the position 68 to 1567 of length 1500 bases.

Primer 3.0

Primer picking results for gi|73858565|ref|NM_000185.3| Homo sapiens serpin peptidase inhibitor, clade D (heparin cofactor), member 1 (SERPIND1), mRNA

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	624	20	60.01	45.00	5.00	2.00	TTGTTAATGCCAGCAGCAAG
RIGHT PRIMER	839	20	59.99	55.00	5.00	2.00	AGGCAGGGTCTGAGAAGTCA

SEQUENCE SIZE: 2237

INCLUDED REGION SIZE: 2237

PRIMER3 predicts the presence of the left and the right primers of length 20 residues in the oligonucleotide query. It gives the presence of the left primer starting from 624th position with GC content of 45% and the right primer at starting from 839th position with GC content of 55%.

Genscan

Sequence /tmp/03_02_13-00:16:12.fasta : 2237 bp : 49.08% C+G : Isochore 2 (43 - 51 C+G%)

Predicted genes/exons:

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr..
1.01	Sngl	+	68	1567	1500	1	0	79	44	1715	0.896	161.64
1.02	PlyA	+	2197	2202	6							1.05

GenScan predicts the presence of the exon in the direct strand from the position 68 to 1567.

1. The probability of the predicted output for the exon is **0.896.**
2. The predicted exon is categorized as a strong exon as the exon score is greater than 100 i.e., **161.64.**
3. Genscan also predicted the peptide of length 499 amino acids coding for the Heparin Cofactor II.

PROTEOMICS

Primary Structure Prediction

Bioedit

Length = 499 amino acids
Molecular Weight = 57067.65 Daltons

Amino Acid		Number	Mol%				
Ala	A	24	4.81	Asn	N	33	6.61
Cys	C	3	0.60	Pro	P	17	3.41
Asp	D	31	6.21	Gln	Q	21	4.21
Glu	E	28	5.61	Arg	R	22	4.41
Phe	F	30	6.01	Ser	S	31	6.21
Gly	G	26	5.21	Thr	T	32	6.41
His	H	14	2.81	Val	V	29	5.81
Ile	I	35	7.01	Trp	W	5	1.00
Lys	K	33	6.61	Tyr	Y	12	2.40
Leu	L	54	10.82				
Met	M	19	3.81				

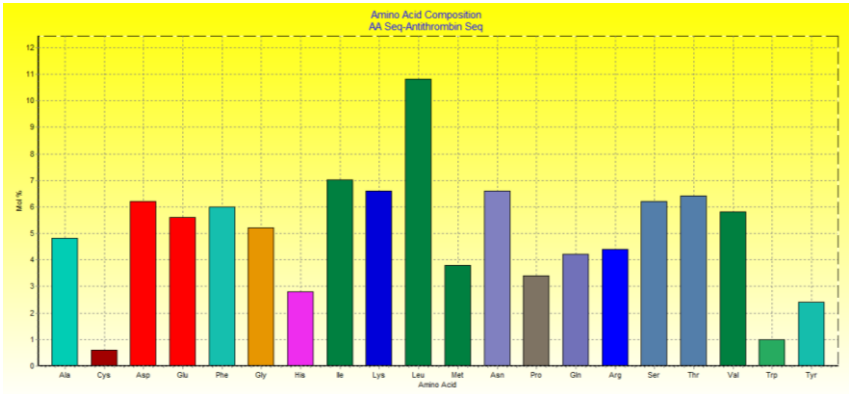

Figure 2 : Aminoacid composition of *Heparin cofactor II (HCII)*

Amino acid composition of Heparin cofactor II (HCII)protein *encoded by SERPIND1* has been predicted by Bioedit & the results reveal that the protein is rich in Leu, Ile, Lys and Asn. The protein has a molecular weight protein of 57067.65 Da containing 499 amino acids.

PROTPARAM

Number of amino acids: 499 **Theoretical pI:** 6.41

Amino acid composition:

Ala (A)	24	4.8%	Phe (F)	30	6.0%
Arg (R)	22	4.4%	Pro (P)	17	3.4%
Asn (N)	33	6.6%	Ser (S)	31	6.2%
Asp (D)	31	6.2%	Thr (T)	32	6.4%
Cys (C)	3	0.6%	Trp (W)	5	1.0%
Gln (Q)	21	4.2%	Tyr (Y)	12	2.4%
Glu (E)	28	5.6%	Val (V)	29	5.8%
Gly (G)	26	5.2%	Pyl (O)	0	0.0%
His (H)	14	2.8%	Sec (U)	0	0.0%
Ile (I)	35	7.0%			
Leu (L)	54	10.8%	(B)	0	0.0%
Lys (K)	33	6.6%	(Z)	0	0.0%
Met (M)	19	3.8%	(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 59

Total number of positively charged residues (Arg + Lys): 55

Formula: C2561H4027N685O747S22

Total number of atoms: 8042

Instability index: 31.68. This classifies the protein as stable.

Aliphatic index: 88.81

Grand average of hydropathicity (GRAVY): -0.236

The protein was slightly acidic in nature, as it has more negatively charged residues (59) than the positively charged ones (55). The protein is classified as stable, as the instability index is computed to be 31.86. The aliphatic index predicts the volume occupied by the aliphatic residue side chains and the index is 88.81. The protein is highly hydrophilic as the Grand Average of Hydropathicity value -0.236, which is very much lesser than 0.05.

SECONDARY STRUCTURE PREDICTION

SOPMA

Sequence length : 499

Alpha helix	(Hh)	:	195	is	39.08%	Beta turn	(Tt) :	27 is 5.41%
3_{10} helix	(Gg)	:	0	is	0.00%	Bend region	(Ss) :	0 is 0.00%
Pi helix	(Ii)	:	0	is	0.00%	Random coil	(Cc) :	203 is 40.68%
Beta bridge	(Bb)	:	0	is	0.00%	Ambigous states	(?) :	0 is 0.00%
Extended strand	(Ee)	:	74	is	14.83%	Other states	:	0 is 0.00%

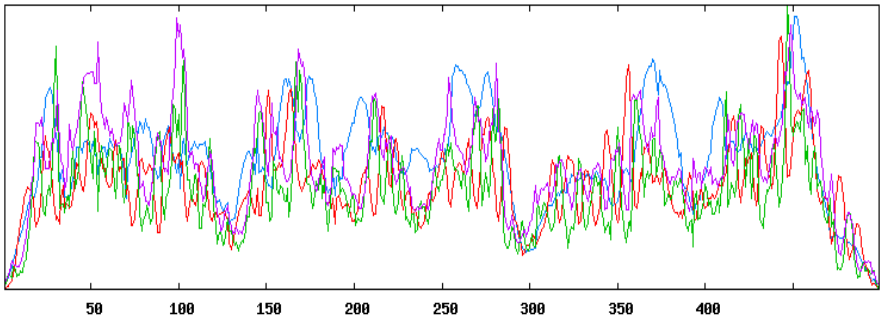

Figure 3: **Secondary structure of Heparin cofactor II (HCII)**

SOPMA (Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments) method was used to predict the secondary structure of Heparin cofactor II (HCII). It was found that 33.08% of amino acids fall in Alpha helix region, 14.83% of amino acid was found to be lays in beta sheet, 5.41% in beta turn and remaining 40.68% tends to form random coil.

POST TRANSLATIONAL MODIFICATIONS

NetNGlyc

Sequence Length: 449

Net N-Glyc results states that in the given protein contains only three N-Glycosylation sites for Threonine at positions 49,188& 387.

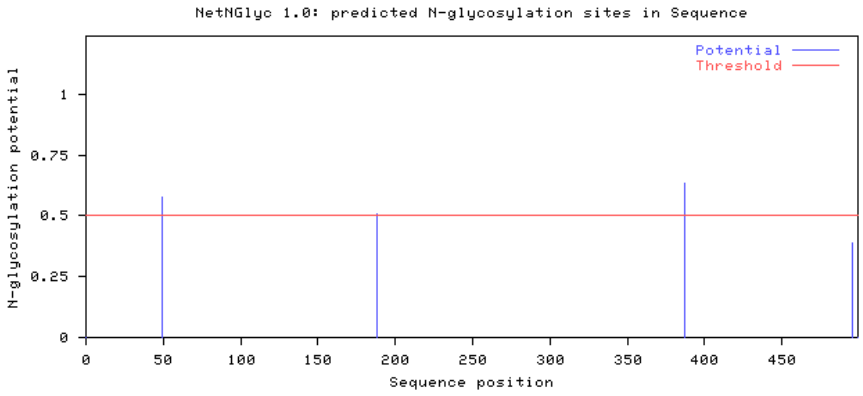

Figure 4 : **Net N-Glycosylation sites of Heparin cofactor II (HCII)**

NetOGlyc

Sequence Length: 449

SERPIND1 coded protein has only one O-Glycosylation unit for Threonine at position 104.

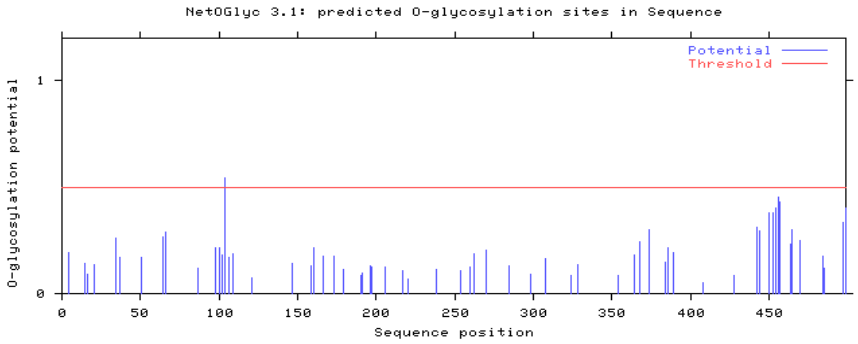

Figure 5 : **Net N-Glycosylation sites of Heparin cofactor II (HCII)**

NetPhos 2.0

Phosphorylation sites predicted: Ser: 9 Thr: 7 Tyr: 4

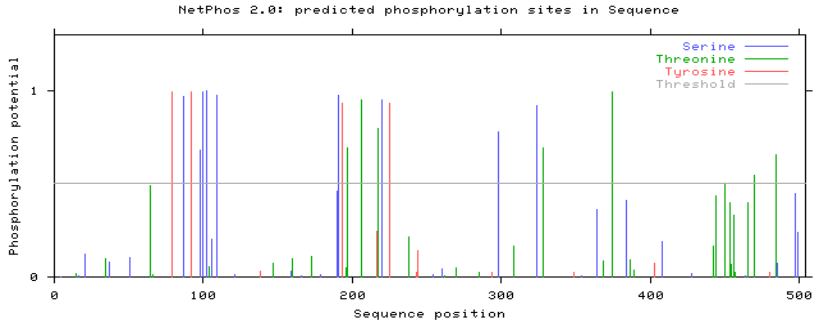

Figure 6 : **Net Phosphorylation sites of Heparin cofactor II (HCII)**

Netphos 2.0 results say that it contains the phosphorylation sites of 9, 7 and 4 for Serine, Threonine & Tyrosine respectively.

CONCLUSION

The official name of SERPIND1 gene is “serpin peptidase inhibitor, clade D (heparin cofactor), member1” which code for Heparin cofactor II (HCII) protein. Defects in SERPIND1 are the cause of thrombophilia due to heparin cofactor 2 deficiency (THPH10), a hemostatic disorder characterized by a tendency to recurrent thrombosis. The SERPIND1 gene of *Homosapiens* codes for a serpin peptidase inhibitor which rapidly inhibits thrombin in the presence of dermatan sulfate or heparin. Genomics states that the pyramidine rich gene codes for 499 aminoacid peptide with the molecular weight of approximately 57067.65 Daltons. The gene contains both left & right primers and several restriction sites. In proteomics primary structure says that it is an acidic, stable protein. Secondary structure is rich in random coils, alpha helix, beta sheets & beta turns. The protein has a total of 20 phosphorylation sites (Serine-9, Threonine-7 & Tyrosine-4) along with three O-Glycosylation

sites and one N-Glycosylation site in its. This insilico studies are very much useful for further advancements in treating thrombiosis.

REFERENCES

1. http://ghr.nlm.nih.gov/gene/SERPIND1
2. R. Carrell and J. Travis. (1985): "α1-Antitrypsin and the serpins: Variation and countervariation". *Trends Biochem. Sci.* **10**: 20–24.
3. Whisstock JC, Silverman GA, Bird PI, Bottomley SP, Kaiserman D, Luke CJ, Pak SC, Reichhart JM, Huntington JA (2010): "Serpins Flex Their Muscle: II. STRUCTURAL INSIGHTS INTO TARGET PEPTIDASE RECOGNITION, POLYMERIZATION, AND TRANSPORT FUNCTIONS". *J Biol Chem* **285** (32): 24307–12.
4. Silverman GA, Whisstock JC, Bottomley SP, Huntington JA, Kaiserman D, Luke CJ, Pak SC, Reichhart JM, Bird PI (2010): "Serpins Flex Their Muscle: I. PUTTING THE CLAMPS ON PROTEOLYSIS IN DIVERSE BIOLOGICAL SYSTEMS". J Biol Chem 285 (32): 24299–305.
5. Hunt LT, Dayhoff MO (1980): "A surprising new protein superfamily containing ovalbumin, antithrombin-III, and α1-proteinase inhibitor". Biochem. Biophys. Res. Commun. 95 (2): 864–71.
6. Irving JA, Pike RN, Lesk AM, Whisstock. (2000:. "Phylogeny of the Serpin Superfamily: Implications of Patterns of Amino Acid Conservation for Structure and Function". Genome Res. 10 (12): 1845–64. doi:10.1101/gr.GR-1478R. PMID 11116082
7. Irving J, Steenbakkers P, Lesk A, Op den Camp H, Pike R, Whisstock J (2002): "Serpins in prokaryotes". Mol Biol Evol 19 (11): 1881–90.
8. Steenbakkers PJ, Irving JA, Harhangi HR, *et al.* (2008): "A serpin in the cellulosome of the anaerobic fungus Piromyces sp. strain E2". *Mycol. Res.* **112** (Pt 8): 999–1006.
9. Rawlings ND, Tolle DP, Barrett AJ. (2004): "Evolutionary families of peptidase inhibitors". Biochem J. 378 (Pt 3): 705–16.
10. EntrezGene:SERPIND1 serpin peptidase inhibitor, clade D (heparin cofactor), member1.
11. C. Geourjon & G. Deleage, 1995, "SOPMA: Significant improvement in protein secondary structure prediction by c prediction from alignments and joint prediction", CABIOS, 11, 681-684.