International Journal of Advanced Research in Engineering and Applied Sciences

ISSN: 2278-6252

AN INSILICO APPROACH TO STUDY AT3 GENE FROM HOMO SAPIENS

Maheswara Reddy Mallu*

Srinivasa Reddy Ronda*

Sandeep Vemula*

Abstract: *Antithrombin is a glycoprotein produced by the liver that inactivates several enzymes of the coagulation system. Antithrombin is also termed Antithrombin III (AT III) and is produced by At3 gene in human beings. The insilico study of At3 gene with genomic and proteomic tools clears that, the At3 gene is pyramidine rich with 14206 bases and 4312931 Daltons (ssDNA) of molecular weight. Primary structure of corresponding Antithrombin III is slightly acidic, stable protein with the instability index of 39.48. Secondary structure defines that it mainly contains the random coils, alpha helix and beta sheets. The gene contains both left & right primers, a hybridization probe and several restriction sites for restriction enzymes. The protein has 28 phosphorylation sites (Serine, Threonine & Tyrosine) along with four N-Glycosylation sites.*

Keywords*: At3 gene, Antithrombin III, Homo sapiens, Primary & Secondary structure, phosphorylation.*

*Department of Biotechnology, K L University, Vaddeswaram, Andhra Pradesh, INDIA.

INTRODUCTION

Antithrombin III (AT III) which is also known as Antithrombin is a protein molecule inactivating several enzymes of the coagulation system in *Homo sapiens.* α-Antithrombin is the dominant form of antithrombin found in blood plasma and has an oligosaccharide occupying each of its four glycosylation sites. A single glycosylation site remains consistently un-occupied in the minor form of antithrombin, β-antithrombin **[1]**. Its activity is increased manyfold by the anticoagulant drug heparin, which enhances the binding of antithrombin to factor II and factor X. Antithrombin has a half-life in blood plasma of around 3 days **[2]**. The normal antithrombin concentration in human blood plasma is high at approximately 0.12 mg/ml, which is equivalent to a molar concentration of 2.3 μM **[3]**. Antithrombin has been isolated from the plasma of a large number of species additional to humans **[4]**. Recombinant antithrombins with properties similar to those of normal human antithrombin have been produced using baculovirus-infected insect cells and mammalian cell lines grown in cell culture **[5,6,7,8]**. Antithrombin begins in its native state, which has a higher free energy compared to the latent state, which it decays to on average after 3 days. The latent state has the same form as the activated state - that is, when it is inhibiting thrombin. As such it is a classic example of the utility of kinetic vs thermodynamic control of protein folding. Antithrombin is used as a protein therapeutic that can be purified from human plasma **[9]** or produced recombinantly (for example, Atryn, which is produced in the milk of goats) **[10,11]**. Antithrombin is approved by the FDA as an anticoagulant for the prevention of clots before, during, or after surgery or birthing in patients with hereditary antithrombin deficiency **[9,11]**. Native antithrombin can be converted to latent antithrombin (L-antithrombin) by heating alone or heating in the presence of citrate **[12,13]**. However, without extreme heating and at 37°C (body temperature) 10% of all antithrombin circulating in the blood is converted to the L-antithrombin over a 24 hour period **[14,15]**.

METHODOLOGY

Many number of genomic & proteomic offline and online tools were used to explore the *At3* gene. The nucleic acid sequence of *Ayt3* gene of *Homo sapiens* was retrieved from NCBI (gi|28906|emb|X68793.1|). Genomic studies were carried out by Bioedit, ORF finder, Primer 3.0 and Genscan followed by proteomic studies to predict primary structure with

Bioedit, Protparam then secondary structure was predicted by GOR4 **[16]**. Post translational modifications are studied with NetOGlyc, NetNGlyc and NetPhos 2.0 tools.

RESULTS AND DISCUSSION

GENOMICS

Bioedit

DNA molecule: gi|28906|emb|X68793.1| H.sapiens gene for antithrombin III

Length = 14206 base pairs

Molecular Weight = 4312931.00 Daltons, single stranded

Molecular Weight = 8626547.00 Daltons, double stranded

G+C content = 46.14% A+T content = 53.86%

Nucleotide	Number	Mol%
A	3715	26.15
C	3338	23.50
G	3217	22.65
T	3936	27.71

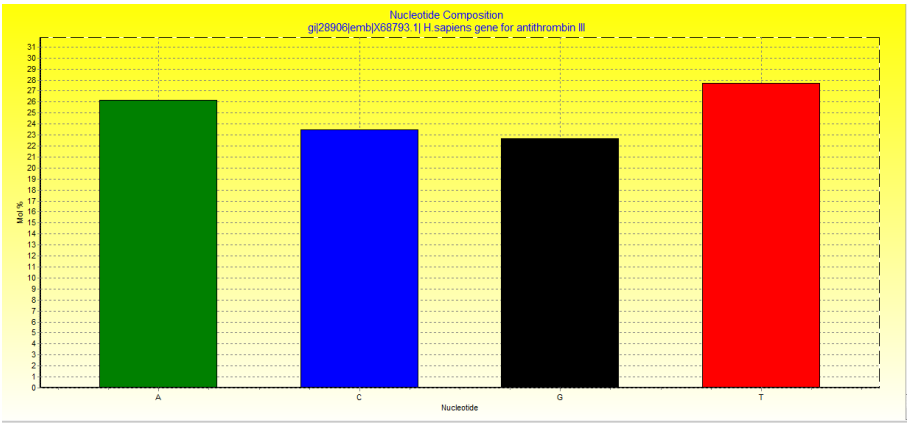

Figure 1 : Nucleotide composition of *At3* gene

Nucleotide composition of At3 gene from *H.sapiens* has been predicted by Bioedit & the results reveals that the sequence is pyramidine rich with the molecular weight of more than 4312931Da (ssDNA) containing 14206 base pairs. The CG & AT content was found to be **46.14%** and **53.86%** respectively. The gene has a total of 89 unique restriction sites specific for the restriction enzymes (W.R.To Bangalore Genei)

ORF Finder

Open Reading Frame Finder predicts the presence of the possible protein coding region sequence.

1. It was identified that At3 gene codes for 69 exons present in both the + and – strands.
2. The largest protein-coding region (exon) was identified in the 2nd frame of the direct strand from the position 3047 to 3460 of length 414 bases.

Primer 3.0

PRIMER PICKING RESULTS FOR gi|28906|emb|X68793.1| H.sapiens gene for antithrombin III

OLIGO	start	len	tm	gc%	any	3'	seq
LEFT PRIMER	2797	20	59.97	50.00	4.00	2.00	GCCCCTGTACTTGGTTCAAA
RIGHT PRIMER	2975	20	59.99	45.00	3.00	0.00	CAATGAGCAGCAAGGACAAA
HYB OLIGO	2820	20	59.95	45.00	6.00	6.00	TTTAGCCTTTCTCTTGGCCA

SEQUENCE SIZE: 14206

INCLUDED REGION SIZE: 14206

PRIMER3 predicts the presence of the left and the right primers of length 20 residues in the oligonucleotide query.

1. Left primer starts from 2797th position with GC content of 50% and the right primer starts from 2975th position with GC content of 45%.
2. The hybridization probe started from 2820th position with GC content of 45%.

Genscan

Predicted genes/exons:

Gn.Ex	Type	S	.Begin	...End	.Len	Fr	Ph	I/Ac	Do/T	CodRg	P....	Tscr..
1.01	Init	+	641	681	41	1	2	80	116	9	0.632	2.56
1.02	Intr	+	2980	3346	367	1	1	130	80	445	0.966	43.05
1.03	Intr	+	5879	6094	216	1	0	94	113	318	0.999	33.60
1.04	Intr	+	7000	7137	138	0	0	87	73	187	0.988	17.76
1.05	Intr	+	7948	8338	391	0	1	92	72	634	0.938	56.50
1.06	Intr	+	10372	10436	65	2	2	110	105	11	0.768	3.44
1.07	Term	+	13811	13987	177	1	0	74	38	129	0.629	4.19
1.08	PlyA	+	14042	14047	6							1.05

All the predicted exons are categorized as weak exons as the exon scores are less than 100**.** It also predicted a peptide of about 464 amino acids in length & a conserved Domain of about 1395 basepairs.

PROTEOMICS

Primary Structure Prediction

Bioedit

Protein: /tmp/03_03_13-06:06:13.fasta|GENSCAN_predicted_peptide_1|464_aa

Length = 464 amino acids

Molecular Weight = 52599.65 Daltons

Amino Acid		Number	Mol%
Ala	A	31	6.68
Cys	C	8	1.72
Asp	D	24	5.17
Glu	E	38	8.19
Phe	F	27	5.82
Gly	G	21	4.53
His	H	5	1.08
Ile	I	24	5.17
Lys	K	37	7.97
Leu	L	45	9.70
Met	M	13	2.80
Asn	N	24	5.17
Pro	P	21	4.53
Gln	Q	13	2.80
Arg	R	23	4.96
Ser	S	35	7.54
Thr	T	26	5.60
Val	V	32	6.90
Trp	W	5	1.08
Tyr	Y	12	2.59

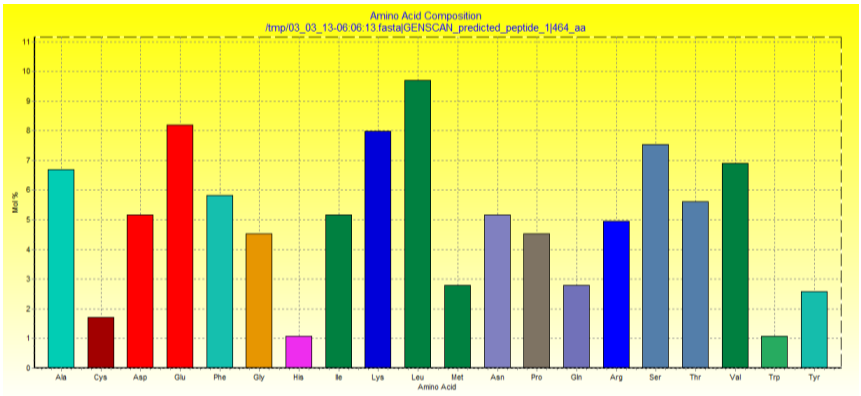

Figure 2 : Aminoacid composition of *Antithrombin III*

Amino acid composition of *Antithrombin III* protein has been predicted by Bioedit & the results reveal that the protein is rich in Leu, Glu and Ser. The protein has a molecular weight protein of 52599.65 Daltons containing 464 amino acids.

PROTPARAM

Number of amino acids: 464

Theoretical pI: 6.32

Amino acid composition

Ala (A)	31	6.7%	Phe (F)	27	5.8%
Arg (R)	23	5.0%	Pro (P)	21	4.5%
Asn (N)	24	5.2%	Ser (S)	35	7.5%
Asp (D)	24	5.2%	Thr (T)	26	5.6%
Cys (C)	8	1.7%	Trp (W)	5	1.1%
Gln (Q)	13	2.8%	Tyr (Y)	12	2.6%
Glu (E)	38	8.2%	Val (V)	32	6.9%
Gly (G)	21	4.5%	Pyl (O)	0	0.0%
His (H)	5	1.1%	Sec (U)	0	0.0%
Ile (I)	24	5.2%			
Leu (L)	45	9.7%	(B)	0	0.0%
Lys (K)	37	8.0%	(Z)	0	0.0%
Met (M)	13	2.8%	(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 62

Total number of positively charged residues (Arg + Lys): 60

Formula: C2355H3718N622O699S21

Total number of atoms: 7415

Instability index:

The instability index (II) is computed to be 39.48

This classifies the protein as stable.

Aliphatic index: 84.68

Grand average of hydropathicity (GRAVY): -0.278

The protein is slightly acidic in nature, as it has more negatively charged residues (62) than the positively charged ones (60). The protein is classified as stable, as the instability index is computed to be 34.48. The aliphatic index predicts the volume occupied by the aliphatic residue side chains and the index is 88.68. The protein is highly hydrophilic as the Grand Average of Hydropathicity value is –0.278, which is very much lesser than 0.05.

PREDICTION OF SECONDARY STRUCTURE

GOR 4

Sequence length : 464

Alpha helix	(Hh)	:	185 is	39.87%
3_{10} helix	(Gg)	:	0 is	0.00%
Pi helix	(Ii)	:	0 is	0.00%
Beta bridge	(Bb)	:	0 is	0.00%
Extended strand	(Ee)	:	71 is	15.30%
Beta turn	(Tt)	:	0 is	0.00%
Bend region	(Ss)	:	0 is	0.00%
Random coil	(Cc)	:	208 is	44.83%
Ambigous states	(?)	:	0 is	0.00%
Other states		:	0 is	0.00%

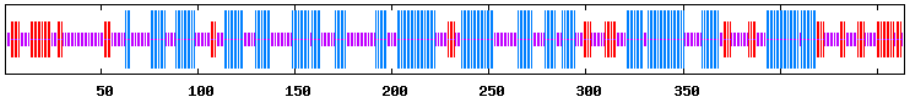

Figure 3: Secondary structure of peptide encode by *At3* gene

GOR4 (Garnier Method) methods was used to predict the secondary structure of Antithrombin III protein. It was found that 39.87% of amino acids fall in alpha helix region, 15.30% of amino acid was found to be lays in beta sheet remaining 44.83% tends to form random coil.

POST TRANSLATIONAL MODIFICATIONS

NetOGlyc

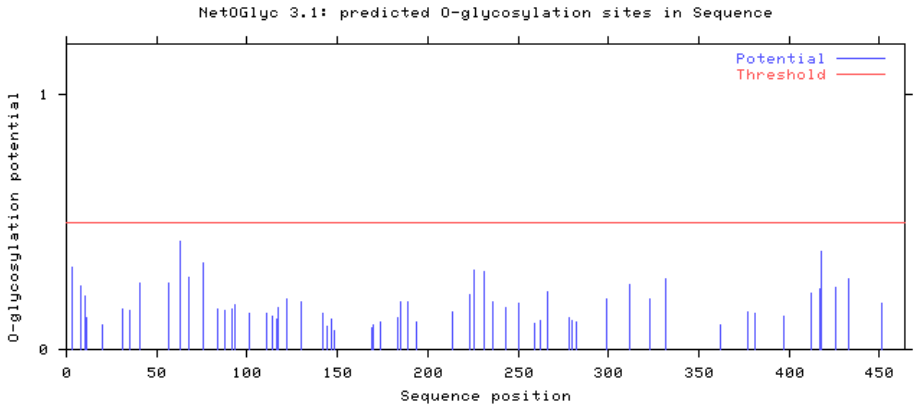

Figure 4 : Net O Glycosylation sites of protein

NetOGlyc results states that in the given protein contains no O-Glycosylation site for Threonine.

NetNGlyc

Asn-Xaa-Ser/Thr sequons in the sequence output below are highlighted in **blue**.
Asparagines predicted to be N-glycosylated are highlighted in **red.**

```
Name:   Sequence          Length:   464
MYSNVIGTVTSGKRKVYLLSLLLIGFWDCVTCHGSPVDICTAKPRDIPMNPMCIYRSPEKKATEDEGSEQKIPEATNRRV     80
WELSKANSRFATTFYQHLADSKNDNDNIFLSPLSISTAFAMTKLGACNDTLQQLMEVFKFDTISEKTSDQIHFFFAKLNC    160
RLYRKANKSSKLVSANRLFGDKSLTFNETYQDISELVYGAKLQPLDFKENAEQSRAAINKWVSNKTEGRITDVIPSEAIN    240
ELTVLVLVNTIYFKGLWKSKFSPENTRKELFYKADGESCSASMMYQEGKFRYRRVAEGTQVLELPFKGDDITMVLILPKP    320
EKSLAKVEKELTPEVLQEWLDELEEMMLVVHMPRFRIEDGFSLKEQLQDMGLVDLFSPEKSKLPGIVAEGRDDLYVSDAF    400
HKAFLEVNEEGSEAAASTAVVIAGRSLNPNRVTFKANRPFLVFIREVPLNTIIFMGRVANPCVK
...n.....................................................n.........................n...     80
.........................n.n...................N.............................n.     160
......N..................N.....................n........n....N..............n     240
........n...............n.....................................................     320
.....................................................................................     400
.......n..................n.n..................n..............                    480
```

(Threshold=0.5)

Antithrombin III protein has four N-Glycosylation Units at the positions 128,167,187 & 224.

NetPhos 2.0

Phosphorylation sites predicted: Ser: 17 Thr: 7 Tyr: 4

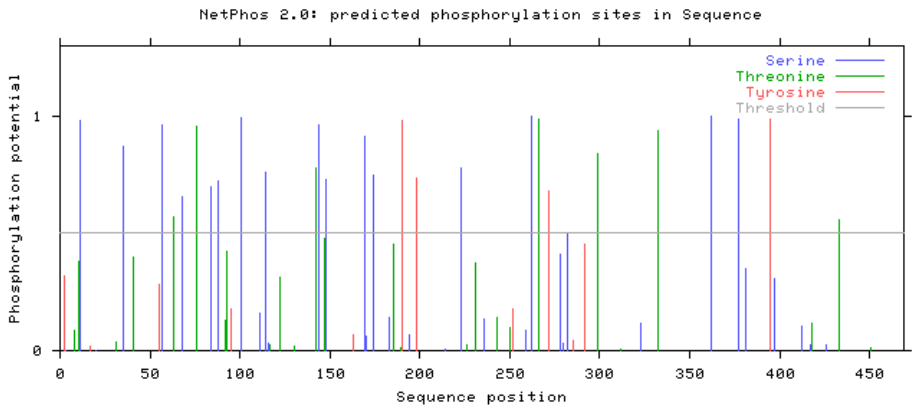

Figure 5 : Net Phosphorylation sites of protein

A Netphos result says that it contains 17, 7 & 4 phosphorylation sites for Serine, Threonine & Tyrosine respectively.

CONCLUSION

Antithrombin is also termed Antithrombin III (AT III) and is produced by At3 gene in human beings. The At3 gene of *Homo sapiens* (gi|28906|emb|X68793.1|) codes for Antithrombin III protein. The gene studies reveal that the pyramidine rich gene codes for 464 aminoacid peptide with the molecular weight of approximately 52599.65 Daltons. The gene contains both left & right primers, a hybridization probe and several restriction sites for restriction

sites. Primary structure confirms it as an acidic, stable protein. Secondary structure reveals that it mainly contains the random coils, alpha helix and beta shttes. The protein has a total of 28 phosphorylation sites (Serine-17, Threonine-7 & Tyrosine-4) along with four N-Glycosylation sites and no O-Glycosylation site. This insilico studies are very much useful for modifying the At3 gene of *Homo sapiens* for further improvements of Antithrombin.

REFERENCES

1. Bjork, I; Olson, JE (1997): *Antithrombin, A bloody important serpin (in Chemistry and Biology of Serpins)*. Plenum Press. pp. 17–33.
2. Collen DJ, Schetz F. *et al.* (1977): "Metabolism of antithrombin III (heparin cofactor) in man: Effects of venous thrombosis of heparin administration". *Eur. J. Clin. Invest* 7 (1): 27–35.
3. Conrad J, Brosstad M. *et al.* (1983): "Molar antithrombin concentration in normal human plasma". *Haemostasis* 13 (6): 363–368.
4. Jordan RE. (1983): "Antithrombin in vertebrate species: Conservation of the heparin-dependent anticoagulant mechanism". *Arch. Biochem. Biophys* 227 (2): 587–595.
5. Stephens AW, Siddiqui A. and Hirs CH. (1987): "Expression of functionally active human antithrombin III". *Proc. Natl. Acad. Sci. USA.* 84 (11): 3886–3890.
6. Zettlmeissl G, Conradt HS. *et al.* (1989): "Characterization of recombinant human antithrombin III synthesized in Chinese hamster ovary cells". *J. Biol. Chem.* 264 (35): 21153–21159.
7. Gillespie LS, Hillesland KK and Knauer DJ. (1991): "Expression of biologically active human antithrombin III by recombinant baculovirus in Spodoptera frugiperda cells". *J. Biol. Chem.* 266 (6): 3995–4001.
8. Ersdal-Badju E, Lu A. *et al.* (1995): "Elimination of glycosylation heterogeneity affecting heparin affinity of recombinant human antithrombin III by expression of a beta-like variant in baculovirus-infected insect cells". *Biochem. J.* 310: 323–330.
9. Thrombate III label
10. FDA website for ATryn (BL 125284)
11. http://www.fda.gov/downloads/BiologicsBloodVaccines/BloodBloodProducts/ApprovedProducts/LicensedProductsBLAs/FractionatedPlasmaProducts/UCM134045.pdf Antithrombin (Recombinant) US Package Insert ATryn for Injection February 3, 2009

12. Chang WS. and Harper PL. (1997): "Commercial antithrombin concentrate contains inactive L-forms of antithrombin". *Thromb. Haemost.* 77 (2): 323–328.
13. Wardell MR, Chang WS. *et al.* (1997): "Preparative induction and characterization of L-antithrombin: a structural homologue of latent plasminogen activator inhibitor-1". *Biochemistry* 36 (42): 13133–13142.
14. Carrell RW, Huntington JA. *et al.* (2001): "The conformational basis of thrombosis". *Thromb. Haemost.* 86 (1): 14–22.
15. Zhou A, Huntington JA. and Carrell RW. (1999): "Formation of the antithrombin heterodimer in vivo and the onset of thrombosis". *Blood* 94 (10): 3388–3396.
16. Geourjon C, Deléage G, SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, PMID: 8808585, Medline.