International Journal of Animal Biotechnology
E-ISSN 2277–4122

http://inpressco.com/category/httpinpressco-comcategoryijab/

Research Article

Proteomic Analysis and Characterization of Amylin (IAPP) in *Homo Sapiens*

Swetha R, Mahesh Pattabhiramaiah*, Reddy M.S and Mandeep Singh

Centre for Applied Genetics, Department of Zoology, Jnanabharathi, Bangalore University, Bangalore-560056, India

Accepted 20 Feb 2016, Available online 26 March 2016, **Vol.6 (2016)**

Abstract

Amylin which is also called as Islet amyloids polypeptide (IAPP), is a pancreatic B-cell hormone co-released with insulin. Amyloids are insoluble fiber and the abnormal accumulation of these fibers in the tissue of the pancreas leads to type II diabetes. Bioinformatics tools were applied to process, prioritise and annotate the raw sequence which is then converted into meaningful information about the proteins. These methods are useful in the medical research as they increase the annotation of proteome through functional and structural genomic efforts. Firstly, the result of primary structure analysis reveals that most of the amylin are hydrophobic in nature due to presence of high non-polar residues content. Amylin has both acidic and basic property as its isoelectric point ranged from 5.31-9.83. The aliphatic index computed by Expasy's Protparam infers that most of the amylin are stable at wide range of temperature (66.84-108.43). The secondary structure analysis by SOPMA reveals that amylin contain more random coils and beta sheets. SMART values revealed an expected range of 0.0000198-27. The Ligand binding site of an amylin was obtained by SWISS-MODEL as it determines the 3D structure of a protein. The RNA structure was predicted by using Genebee service software.

Keywords: *IAPP, Type 2 Diabetes, Expasy ProtParam, SOPMA, SMART, SWISS-MODEL.*

Introduction

Amylin is a protein which is insoluble in low tissue concentration. The formation of amyloids, is known to cause non-neuronal diseases such as Diabetes mellitus type II. Amylin forms long stable fibres which are arranged in regular pattern [Nanga *et al*, 2008]. Islet amyloid polypeptide (IAPP, also called amylin) is associated with type II diabetes, a disease that has maximized over the past two decades and now afflicts an estimated 20 million people in America and 100 million worldwide [Hossain *et al*, 2007]. The role of IAPP in diabetes is uncertain [Hoppener *et al,* 2002], but there is escalating evidence for the consequence of amyloid formation associated with this disease. Ninety-five percent of analysed individuals stain positive posthumous for pancreatic amyloid deposits composed of mature, fibrillar IAPP [Hoppener *et al,* 2000]. Gene responsible for the formation of amyloid is located on the 12p12.1 chromosome. Pancreatic extracts from respective person with type II diabetes have been found to contain extracellular deposits of the 37-residue peptide hormone IAPP that is expressed, stored, and secreted from the beta-islet cells of the pancreas [Cooper, 1994; Cooper *et al,* 1987]. IAPP was first identified in 1987, and was considered as a Potential target for the treatment of diabetes [Schmitz *et al,* 2004; Yan *et al,* 2006]. Amyloids are insoluble fibrillar aggregates of proteins / peptides. Type-2 diabetes mellitus (DM-2) was one such disease associated with human islet amyloid polypeptide (hIAPP, also called amylin), which is a peptide of 37 residues secreted from beta-cells of the pancreas. hIAPP (Human Islet Amyloid Polypeptide) is monomeric in its physiological state but aggregated in the disease state. Evidence from the cell and animal studies indicates a link between hIAPP misfolding and pancreatic beta cell dysfunction. [Balali-Mood *et al,* 2005; Cooper *et al,* 1987; Eisenhaber *et al,* 1996; Gill and Von Hippel, 1989]. The misfolding or abnormal accumulation of these peptides in the tissue of the pancreas leads to the metabolic disorder DM-2 [Zanuy *et al,* 2003]. Misfolding or unfolding of native protein exposes hydrophobic regions which results in unstable intermediates of revised conformation that have the propensity to form oligomers, where these oligomers form pathogenic subunits and crossed beta sheets. The soluble oligomers become toxic and accumulate to promote apoptosis [Balali-Mood *et al,* 2005].

Researchers are trying to find out the aberrant metabolic conditions and factors which can possibly trigger or can have the potential to hinder amyloid formation. Some evidence suggests that altered hIAPP

*Corresponding author: **Mahesh Pattabhiramaiah**

sequences and abnormal metabolic conditions might be responsible for development of this disease [Hayden, 2002]. Human amylin exhibits physicochemical properties predisposing the peptide hormone to accumulate and form Amyloids fibers, which may play a part in B-cell destruction in type 2 diabetes.

Several peptide segments derived from the amino acid sequence of IAPP are capable of composing amyloid-like fibrils, with earlier work suggesting that the segment containing residues 22–29 forms the spine of the fibril under physiological conditions [Westermark *et al,* 1990; Moriarty and Raleigh, 1999; Goldsbury *et al,* 2000; Tenidis *et al,* 2000.]. Amylin is a pancreatic B-cell hormone co-released with insulin in response to food intake [Young and Denaro, 1998].

The wealth of amylin protein sequence information that has been made publicly available in recent years requires the advancement of high-throughput functional genomics and proteomics approaches for its analysis. Such approaches need suitable data assimilation procedures and a high level of annotation in order to gain maximum benefit from the results generated. Identification of proteins of interest from a particular biological study requires the application of bioinformatics tools to process and compute the data. From a protein function standpoint, transfer of annotation from known proteins to a novel destination is currently the only practical way to convert vast quantities of raw sequence data into meaningful information. Advanced bioinformatics tools now provide more sophisticated methods to transfer functional annotation, consolidating sequence, family profile and structural search methodology. The importance of these approaches to medical research is growing as we move to annotate the proteome through functional and structural genomic efforts.

Detailed knowledge of amylin and their properties can be extensively studied through the physiochemical and the structural properties of the proteins are well figured out with the use of computational tools. The statistics about the sequence of amylin protein such as number of amino acid, frequency is predicted by CLC work bench (http:// www.clc.bio. com/index. php?id=28). Modular Architecture Research Tool (SMART) is a biological database that is used in the identification and analysis of protein domains within protein sequences [Letunic *et al,* 2009; Schultz *et al,* 1998]. Sequence length, and the physico-chemical properties of a amylin protein such as molecular weight, atomic composition, extinction coefficient, GRAVY, aliphatic index, instability index, etc. can be computed by ProtParam. The TMpred program makes a prediction of membrane-spanning regions and their orientation. The algorithm was based on the statistical analysis of TMbase, a database of naturally occurring trans-membrane proteins. The prediction was made using a combination of several weight-matrices for scoring [Hofmann and Stoffel, 1993]. MUSCLE stands for MUltiple Sequence Comparison by Log-Expectation. MUSCLE is claimed to achieve both better average accuracy and a better speed than ClustalW2 or T-Coffee. The protein 3D model and its characteristics can be predicted by Swiss model server [Tsetlin and Hucho, 2004]. Reverse Translate accepts a protein sequence as input and uses a codon usage table to achieve a DNA sequence representing the most likely non-degenerate coding sequence. A consent sequence derived from all the possible codons for each amino acid is also returned. Further Computer-aided techniques for the productive identification and optimization of novel molecules with a desired biological activity have become a part of the drug discovery process.

The raw sequence information of all the proteins and nucleic acid can convert to analytical and relative information with the help of soft computing tools [Ashokan and Pillai, 2008]. Prediction of protein function is important application of bioinformatics [Prashanth *et al,* 2010]. The amino acid sequence provides most of the information required for determining and characterizing the molecule's function, physical and chemical properties. Sequence analysis and physicochemical characterization of proteins using biocomputation tool have been done by many researchers and reported. [Ashokan *et al,* 2011; Madhu and Mahesh, 2015; Mahesh *et al,* 2015; Praveen and Mahesh, 2015; Vishwanath *et al,* 2015; Mandeep Singh *et al,* 2016; Amrutha *et al,* 2015].

The main objective of this study is to determine the physiochemical characterization of the amyloid protein and to predict its function. As Amyloids are associated with type 2 diabetes, characterization of Amylin can be used in the treatment of type 2 diabetes.

Materials and methods

Protein sequence retrieval: The Protein Sequences of amylin (10 sequences) were retrieved in FASTA format from NCBI database (Table 1).

Amino acid Composition: The amino acid composition of selected proteins were computed using the tool CLC free workbench (www.clc.bio.com/..../clc-main-workbench), tabulated in (Table 2).

Primary structure analysis: Counts of hydrophobic and hydrophilic residues were calculated from the primary structure analysis by CLC workbench (Table 3).

Physico-chemical parameters: The physicochemical parameters such as theoretical isoelectric point (Ip), molecular weight, total number of positive and negative residues, extinction coefficient, instability index [Gill and Von Hippel, 1989] aliphatic index [Eisenhaber *et al*, 1996] and grand average hydropathy (GRAVY) [Kitchen *et al*, 2007] were computed using the Expasy's ProtParam server [Mugilan *et al*, 2010], and tabulated in (Table 4).

Secondary structure prediction: The secondary structure was predicted by self-optimized prediction method with alignment by SOPMA server (http://jing.cz3.nus. edu.sg/cgi-bin/ svmprot.cgi.) [Ashokan *et al*, 2011] (Table 5).

Table1: Selected AMYLIN protein sequences retrieved from NCBI

Sl No	Species	ID	Length	Protein Sequence
1	*Homo sapiens*	AAA35524.1	62aa	HQVEKRKCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTYGKRNAVEV LKREPLNYLPL
2	*Homo sapiens*	NP_000406.1	89aa	MGILKLQVFLIVLSVALNHLKATPIESHQVEKRKCNTATCATQRLANFLVHS SNNFGAILSSTNVGSNTYGKRNAVEVLKREPLNYLPL
3	*Homo sapiens*	3HGZ|E	37aa	KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY
4	*Homo sapiens*	2KB8|A	37aa	KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTX
5	*Homo sapiens*	2L86|A	38aa	KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTYX
6	*Homo sapiens*	AAC50301.1	427aa	MMDAQYKCYDRMQQLPAYQGEGPYCNRTWDGWLCWDDTPAGVLSYQFC PDYFPDFDPSEKVTKYCGVWFKHPENNRTWSNYTMCNAFTPEKLKNAYV LYYLAIVGHSLSIFTLVISLGIFVFFRSLGCQRVTLHKNMFLTYILNSMIIIIHLV EVVPNGELVRRDPVSCKILHFFHQYMMACNYFWMLCEGIYLHTLIVVAVFT EKQRLRWYYLLGWGFPLVPTTIHAITRAVYFNDNCWLSVETHLLYIIHGPV MAALVVNFFFLLNIVRVLVTKMRETHEAESHMYLKAVKATMTLVPLLGIQF VVFPWRPSNKMLGKIYDYVMHSLIHFQGFFVATIYCFCNNEVQTTVKRQW AQFKIQWNQRWGRRPSNRSARAAAAAAAEAGDIPIYICHQELRNEPANNQG EESAEIIPLNIIEQESSA
7	*Homo sapiens*	AAC50300.1	474aa	YCNRTWDGWLCWDDTPAGVLSYQFCPDYFPDFDPSEKVTKYCDEKGVWF KHPENNRTWSNYTMCNAFTPEKLKNAYVLYYLAIVGHSLSIFTLVISLGIFVF FRSLGCQRVTLHKNMFLTYILNSMIIIIHLVEVVPNGELVRRDPVSCKILHFF HQYMMACNYFWMLCEGIYLHTLIVVAVFTEKQRLRWYYLLGWGFPLVPTT IHAITRAVYFNDNCWLSVETHLLYIIHGPVMAALVVNFFFLLNIVRVLVTKM RETHEAESHMYLKAVKATMILVPLLGIQFVVFPWRPSNKMLGKIYDYVMHS LIHFQGFFVATIYCFCNNEVQTTVKRQWAQFKIQWNQRWGRRPSNRSARA AAAAAEAGDIPIYICHQEPRNEPANNQGEESAEIIPLNIIEQESSA
8	*Homo sapiens*	NP_001733.1	474aa	MRFTFTSRCLALFLLLNHPTPILPAFSNQTYPTIEPKPFLYVVGRKKMMDAQ YKCYDRMQQLPAYQGEGPYCNRTWDGWLCWDDTPAGVLSYQFCPDYFPD FDPSEKVTKYCDEKGVWFKHPENNRTWSNYTMCNAFTPEKLKNAYVLYYL AIVGHSLSIFTLVISLGIFVFFRSLGCQRVTLHKNMFLTYILNSMIIIIHLVEVVP NGELVRRDPVSCKILHFFHQYMMACNYFWMLCEGIYLHTLIVVAVFTEKQR LRWYYLLGWGFPLVPTTIHAITRAVYFNDNCWLSVETHLLYIIHGPVMAAL VVNFFFLLNIVRVLVTKMRETHEAESHMYLKAVKATMILVPLLGIQFVVFP WRPSNKMLGKIYDYVMHSLIHFQGFFVATIYCFCNNEVQTTVKRQWAQFKI QWNQRWGRRPSNRSARAAAAAAEAGDIPIYICHQELRNEPANNQGEESAEI IPLNIIEQESSA
9	*Homo sapiens*	2WK3|D	42aa	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVVIA
10	*Homo sapiens*	2LMP|K	40aa	DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGGVV

Table 2: Representation of frequency of amino acids in amylin

Amino acid	AAA35524.1	NP_000406.1	3HGZ|E	2KB8|A	2L86|A	AAC50301.1	AAC50300.1	NP_001733.1	2WK3|D	2LMP|K
Alanine (A)	0.081	0.079	0.108	0.108	0.105	0.066	0.063	0.063	0.095	0.075
Cysteine (C)	0.032	0.022	0.054	0.054	0.053	0.033	0.032	0.032	0	0
Aspartic Acid (D)	0	0	0	0	0	0.03	0.027	0.027	0.071	0.075
Glutamic Acid (E)	0.048	0.045	0	0	0	0.052	0.049	0.049	0.071	0.075
Phenylalanine (F)	0.032	0.034	0.054	0.054	0.053	0.061	0.065	0.065	0.071	0.075
Glycine (G)	0.048	0.045	0.054	0.054	0.053	0.044	0.042	0.042	0.143	0.15
Histidine (H)	0.032	0.034	0.027	0.027	0.026	0.035	0.034	0.034	0.071	0.075
Isoleucine (I)	0.016	0.045	0.027	0.027	0.026	0.068	0.068	0.068	0.071	0.05
Lysine (K)	0.065	0.067	0.027	0.027	0.026	0.04	0.042	0.042	0.048	0.05
Leucine (L)	0.113	0.146	0.081	0.081	0.079	0.089	0.093	0.095	0.048	0.05
Methionine (M)	0	0.011	0	0	0	0.035	0.034	0.034	0.024	0.025
Asparagine (N)	0.129	0.101	0.162	0.162	0.158	0.054	0.053	0.053	0.024	0.025
Proline (P)	0.032	0.034	0	0	0	0.047	0.057	0.055	0	0
Glutamine (Q)	0.032	0.034	0.027	0.027	0.026	0.042	0.04	0.04	0.024	0.025
Arginine (R)	0.065	0.045	0.027	0.027	0.026	0.047	0.049	0.049	0.024	0.025
Serine (S)	0.081	0.079	0.135	0.135	0.132	0.042	0.042	0.042	0.048	0.05
Threonine (T)	0.081	0.067	0.135	0.135	0.132	0.052	0.055	0.055	0	0
Valine (V)	0.081	0.09	0.054	0.054	0.053	0.077	0.074	0.074	0.143	0.15
Tryptophan (W)	0	0	0	0	0	0.03	0.027	0.027	0	0
Tyrosine (Y)	0.032	0.022	0.027	0	0.026	0.056	0.055	0.055	0.024	0.025

Table 3: Hydrophobic and Hydrophilic residues content computed by CLC Workbench

ID number	AAA35524.1	NP_000406.1	3HGZ|E	2KB8|A	2L86|A	AAC50301.1	AAC50300.1	NP_001733.1	2WK3|D	2LMP|K
Counts of Hydrophobic residues(A,F,G,I,L,M,P,V,W)	25	43	14	14	14	221	248	248	25	23
Counts of Hydrophilic residues(C,N,Q,S,T,Y)	24	29	20	19	20	119	131	131	5	5

Table 4: Parameters computed by Expasy ProtParam

ID no	PI	Mol wt	-R	+R	EC	II	AI	GRAVY
AAA35524.1	9.78	6875.8	3	8	3105	23.85	81.77	-0.456
NP_000406.1	9.83	9806.4	4	10	3105	26.96	108.43	0.035
3HGZ|E	8.9	3906.3	0	2	1615	1.94	68.65	-0.097
2KB8|A	8.96	3854.4	0	2	125	1.67	68.65	-0.062
2L86|A	8.9	4017.6	0	2	1615	1.89	66.84	-0.095
AAC50301.1	7.88	49882	35	37	108135	42.81	90.16	0.041
AAC50300.1	8.18	47171.9	32	35	103540	43.84	93.84	0.105
NP_001733.1	8.6	55344.5	36	43	111115	44.29	91.1	0.061
2WK3|D	5.31	4514.1	6	3	1490	18.17	97.38	0.205
2LMP|K	5.31	4329.8	6	3	1490	18.58	90	0.057

Table 5: Representation of helix, sheet, turn, coils by through online tool SOPMA

	ID no	AAA35524.1	NP_000406.1	3HGZ|E	2KB8|A	2L86|A	AAC50301.1	AAC50300.1	NP_001733.1	2WK3|D	2LMP|K
Helix(H)	Residue totals	22	41	10	10	10	156	141	171	2	2
	Percentage%	35.48	46.07	27.03	26.32	26.32	36.53	34.9	36.08	4.76	5
Sheet(E)	Residue totals	12	15	8	9	9	114	114	125	22	19
	Percentage%	19.35	16.85	21.62	23.68	23.68	26.7	28.22	26.37	52.38	47.5
Turn(T)	Residue totals	2	3	4	4	4	44	41	42	8	7
	Percentage%	3.23	3.37	10.81	10.53	10.53	10.3	10.15	8.86	19.05	17.5
Coils(C)	Residue totals	26	30	15	15	15	113	108	136	10	12
	Percentage%	41.94	33.71	40.54	39.47	39.47	26.46	26.73	28.69	23.81	30

Table 6: SMART analysis of Amylin protein sequences

ID number	Start	End	E –value
AAA35524.1	5	47	5.47e-17
NP_000406.1	32	74	5.47e-17
3HGZ|E	1	37	8.77e-7
2KB8|A	1	36	0.0000198
2L86|A	1	37	8.77e-7
AAC50301.1	21	96	7.49e-27
AAC50300.1	1	73	4.72e-24
NP_001733.1	68	143	7.49e-27
2WK3|D			No domain
2LMP|K			No domain

Table 7: SVMprot analysis of Amylin protein sequences

Accession number	Protein family name																			
	AAA35524.1		NP_000406.1		3HGZ|E		2KB8|A	2L86|A			AAC50301.1		AAC50300.1		NP_001733.1		2WK3|D		2LMP|K	
	R Value	P Value	R Value	P Value	R Value	P Value	R Value	P Value	R Value	P Value	R Value	P Value	R Value	P Value	R Value	P Value	R Value	P Value	R Value	P Value
Iron binding	1	59	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
All DNA binding	1	59	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Photoreceptor	1	59	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Magnesium binding	1	59	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Hormone	NA	NA	6	99	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Transmembrane	NA	NA	1.2	65	NA	NA	NA	NA	NA	NA	6.8	99	10	99	10	99	NA	NA	NA	NA
All lipid binding proteins	NA	NA	1.1	62	NA	NA	NA	NA	NA	NA	2.1	85	2.8	93	2.8	93	NA	NA	NA	NA
Outer membrane	NA	NA	1	59	NA	NA	NA	NA	NA	NA	1	59	NA	NA	NA	NA	NA	NA	NA	NA
Metal binding	NA	NA	1	59	NA	NA	NA	NA	NA	NA	1	59	1.1	62	1.1	62	NA	NA	NA	NA
Calcium binding	NA	NA	1	59	NA	NA	NA	NA	NA	NA	NA	NA	1	59	1	59	NA	NA	NA	NA
TC 1 C channels/pores	NA	NA	1	59	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
G-Protein Coupled Receptors	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	4.4	98	8.5	99	8.5	99	NA	NA	NA	NA
7 Transmembrane receptor	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.2	65	1	59	1	59	NA	NA	NA	NA
TC 1 A channels/pores	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1.1	62	1.2	65	1.2	65	NA	NA	NA	NA
7 Transmembrane receptor (secretin family)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	59	1	59	1	59	NA	NA	NA	NA
Chlorophyll biosynthesis	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	1	59	1	59	NA	NA	NA	NA

(R value and P values are in percentage)

Table 8: RNA structure stems with free energy (ref/NP_000406.1)

Stem no	1	2	3	4	5	6	7	8	9	10	11
Free energy(Kkal/mol)	-15.8	-13	-10	-9.4	-8.7	-8.4	-8.2	-7.7	-6.9	-6.2	-6.2

Domain architecture analysis: Domain organization and domain composition was analyzed using Simple Modular Architecture Research Tool (SMART) (Table 6).

SVMprot analysis: SVMprot tool was used to predict protein function, and to classify the distantly related proteins (Table 7).

Trans-membrane region prediction: TMpred was used to predict transmembrane helices. The TMpred software is available through internet access (http://www.ch.embnet.org/software/TMPRED_form.html) (Fig 1).

Sequence Homology Analysis: The sequence homology was analyzed by MUSCLE (http://www.ebi.ac.uk/ Tools/msa/muscle/) (Fig 2).

RNA structure prediction: The protein sequences were reverse transcribed to DNA using Sequence Manipulation Suite (SMS) (http://www.bioinformatics.org/ sms2/ rev_trans.html). The reverse transcribed DNA was converted to RNA using transcriptional and translational tool (http://www.attotron.com/cybertory/analysis/trans.html). RNA structure was predicted using (http://www.genebee.msu.su/services/rna2_reduced.html) (Fig 3).

Swiss model: Homology-modeling was performed using SWISS-MODEL accessible via the Expasy web server. (http://swissmodel.expasy.org/) (Fig 4)

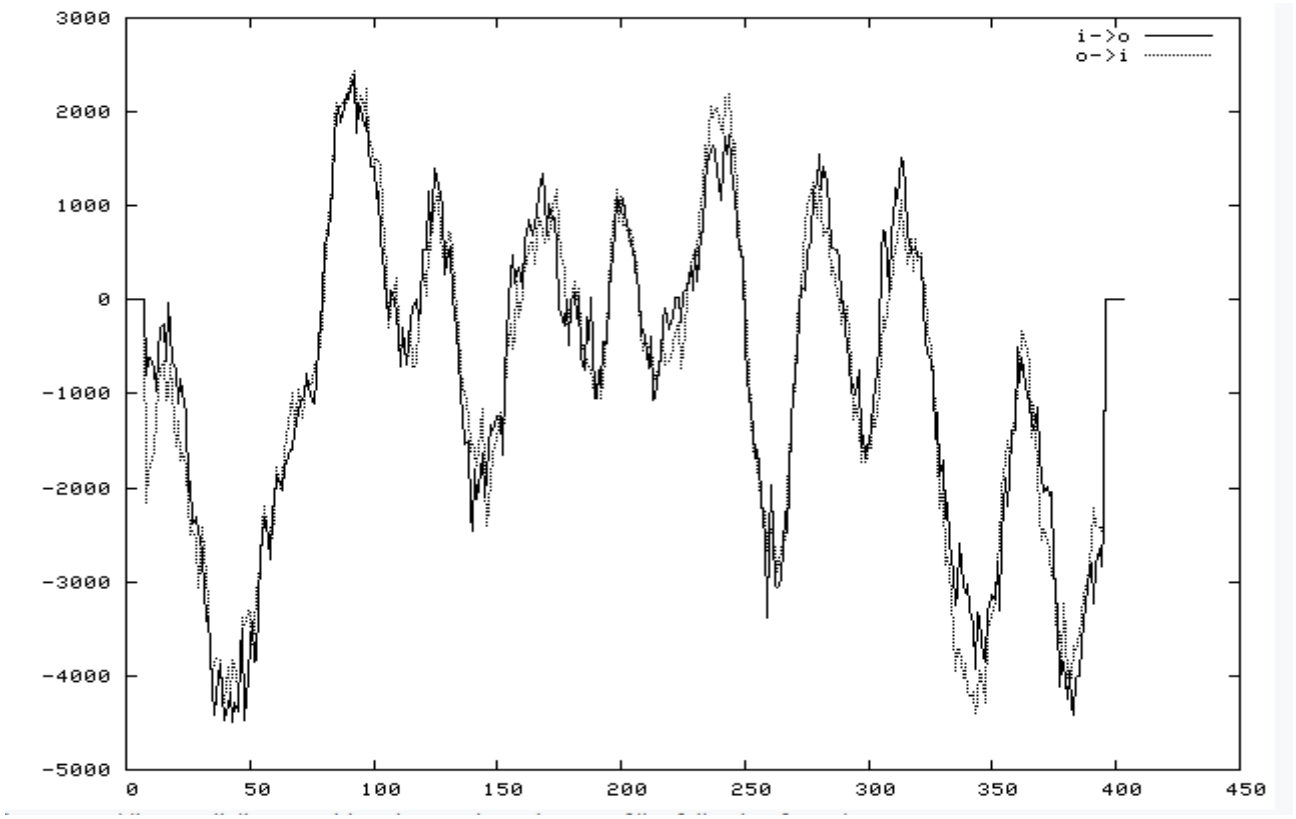


Fig.1 TMPred graph: (ref AAC50300.1)

```
gi|262367769|pdb|2WK3|D          ---------HDSGYEVH---HQKLVFFAE--------------------------------
gi|365812858|pdb|2LMP|K          ---------HDSGYEVH---HQKLVFFAE--------------------------------
gi|1117797|gb|AAC50301.1|calcito WRPSNKMLGKIYDYVMHSLIHFQGFFVATIYCFCNNEVQTTVKRQWAQFKIQWNQRWGRR
gi|1117795|gb|AAC50300.1|        WRPSNKMLGKIYDYVMHSLIHFQGFFVATIYCFCNNEVQTTVKRQWAQFKIQWNQRWGRR
gi|4502547|ref|NP_001733.1|      WRPSNKMLGKIYDYVMHSLIHFQGFFVATIYCFCNNEVQTTVKRQWAQFKIQWNQRWGRR
gi|457131|gb|AAA35524.1|         ---------RLANFLVHSSNNFGAILSST--------------------------------
gi|4557655|ref|NP_000406.1|      ---------RLANFLVHSSNNFGAILSST--------------------------------
gi|339961185|pdb|2L86|A          ---------RLANFLVHSSNNFGAILSST--------------------------------
gi|270346547|pdb|3HGZ|E          ---------RLANFLVHSSNNFGAILSST--------------------------------
gi|224036171|pdb|2KB8|A          ---------RLANFLVHSSNNFGAILSST--------------------------------
                                          .  .: :*   :   .: :
```

Identity (*): Strongly similar (:): Weakly similar (.)

Fig.2 Multiple sequence alignment of Amylin protein sequence by MUSCLE software

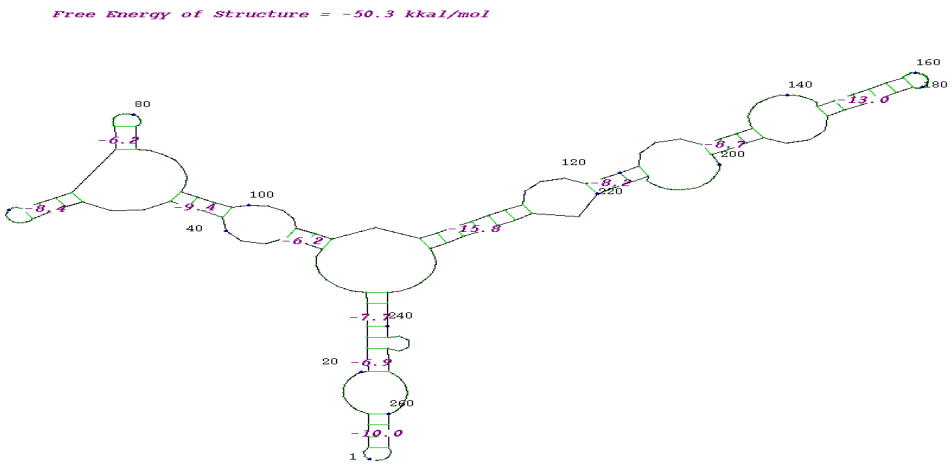
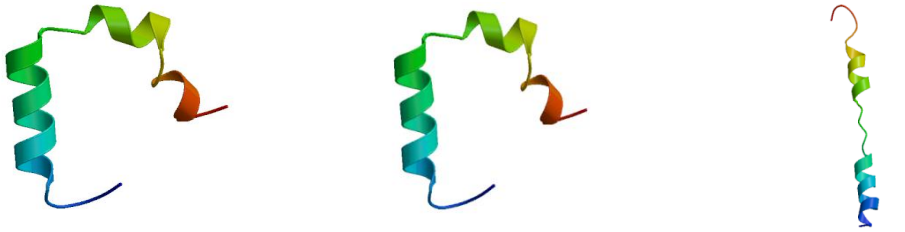


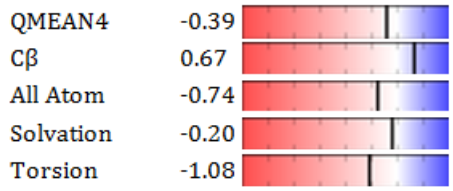
Fig.3 RNA structure prediction: (AAC50300.1)

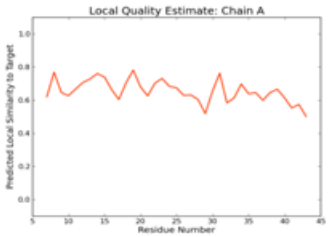


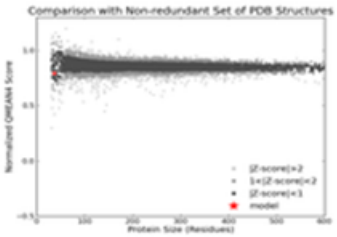
Model 1 Model 2 Model 3

Fig.4 SWISS MODEL: (AAA35524.1)

Model #01	File	Built with	Oligo-State	Ligands	GMQE	QMEAN4
	PDB	ProMod Version 3.70.	MONOMER	None	0.68	-0.39







Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Seq Similarity	Range	Coverage	Description
2l86.1.A	100.00	monomer	BLAST	NMR	NA	0.61	7 - 43	0.60	Islet amyloid polypeptide

The template contained no ligands.

Target HQVEKRKCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTYGKRNAVEVLKREPLNYLPL

2l86.1.A ------KCNTATCATQRLANFLVHSSNNFGAILSSTNVGSNTY----------

Fig.5 Model building using SWISS MODEL (AAA35524.1)

Fig.6 3D view of the structure with ligand binding: (AAA35524.1)

Conclusion

Amylin, a naturally occurring hormone, is a normal product of beta cells and is co-released with insulin.

In the present work, we have addressed the amino acid composition, physiochemical properties, structure and function of amyloid fibrils by using a domain of the diabetes-associated IAPP peptides as model.

Amyloid fibrils formation varies greatly with environmental conditions like temperature, pH, oxidative stress, ionic strength, peptide concentration etc. These highly ordered protein fibrils form aggregates and deposition of which leads to several degenerative diseases such as Alzheimer‟s disease, Parkinson‟s disease, Type II diabetes and so on. The proteomic analysis and characterization of amylin is

done so that a complete knowledge of amylin protein is determined and a site of interest which is to be treated can be identified easily and novel drug can be induced. Usage of bio informatics tool gives a complete knowledge to treat type 2 diabetes which is associated with formation of amyloids from amylin protein.

The understanding of their physiochemical characteristics, function, structure and their mechanism of action can serve as a key therapeutic approach towards treating amyloid diseases.

Acknowledgements

Our sincere thanks to our Co-ordinator, Professor Dr. M. S. Reddy, for the valuable support and encouragement.

References

Abedini.A, Raleigh.D.P, (2005) The role of His-18 in amyloid formation by human islet amyloid polypeptide, *Biochemistry* 44 16284e1629.

Altschul.S.F, Madden.T.L, Schaffer.A.A, Zhang.J, Zhang.Z, Miller.W and Lipman.D.J (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search program's. *Nucleic Acids Res*, 25, 3389-3402.

Amrutha.K.M, Mahesh Pattabhiramaiah and Anusha.P.K (2015) Bio computational Characterization of Wolbachia Surface Protien in Different Species of Drosophila *International Journal of Engineering Research and General Science*. Volume 3; Issue 6; Page no 382-396.

Ashokan.K and Pillai.M. (2008) Characterization of biomimetic material, Mussel adhesive protein, using computer tools and servers. *The Internet Journal of Bioengineering*. Vol 4(1).

Ashokan.K.V, Mundaganur.D.S, and Mundaganur.Y.D. (2011) Catalase: Phylogenetic characterization to explore protein cluster. *Journal of research in Bioinformatics*. 1:001-008.

Balali-Mood, K. R.H. Ashley, T. Hauss, J.P and Bradshaw, (2005) Neutron diffraction reveals sequence-specific membrane insertion of pre-fibrillar islet amyloid polypeptide and inhibition by rifampicin, *FEBS Lett*. 579 1143e1148.

Benkert, P., Biasini, M. & Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics*, 27, 343-350.

Butler, A. E., Jang, J., Gurlo, T., Carty, M. D., Soeller, W. C., and Butler, P. C. (2004) Diabetes due to a progressive defect in -cell mass in rats transgenic for human islet amyloid polypeptide (HIP Rat): a new model for type 2 diabetes. *Diabetes* 53, 1509–1516

Cooper.G.J. (1994) Amylin compared with calcitonin gene-related peptide: Structure, biology, and relevance to metabolic disease. *Endocr. Rev*. 15: 163–201.

Cooper.G.J.S, Willis.A.C, Clark.A, Turner.R.C, Sim.R.B and Reid.K.B.M. (1987) Purification and characterization of a peptide from amyloid-rich pancreases of type 2 diabetic patients. *Proc. Natl. Acad. Sci*. 84: 8628–8632.

Eisenhaber, F., Imperiale, F., Argos, P. & Froemmel, C. (1996) Prediction of secondary structural content of proteins from their amino acid composition alone. I. New analytic vector decomposition methods. *Proteins: Struct. Funct. Design*.25:157-168.

Gill, S.C. & Von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem*.182: 319-326.

Goldsbury.C, Goldie.K, Pellaud.J, Seelig.J, Frey.P, Muller.S.A, Kistler.J, Cooper.G.J.S, and Aebi.U. (2000) Amyloid fibril formation from full-length and fragments of amylin. *J. Struct. Biol.* 130: 352– 362.

Guex, N. & Peitsch, (1997) M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis,* 18, 2714-2723.

Guruprasad, K., Reddy, B.V.P. & Pandit, M.W. (1990) Correlation between stability of a 273 protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence. *Prot.Eng*. 4 (2):155-164.

Hayden, M.R. (2002) Islet amyloid, metabolic syndrome, and the natural progressive history of type 2 diabetes mellitus. J.Pancreas (Online) 3 126e13.

Hoppener.J.W.M, Ahren.B and Lips.C.J.M (2000) Islet amyloid and type 2 diabetes mellitus. *N. Engl. J. Med*. 343: 411–419.

Hoppener.J.W.M, Nieuwenhuis.M.G, Vroom.T.M, Ahren.B and Lips.C.J.M. (2002) Role of islet amyloid in type 2 diabetes mellitus: Consequence or cause? *Mol. Cell. Endocrinol*. 197: 205–212.

Hofmann, K. and Stoffel, W. (1993) TMbase – A database of membrane spanning proteins segments Biol. Chem. *Hoppe-Seyler* **374**,166.

Hossain.P, Kawar.B and El Nahas.M (2007) Obesity and diabetes in the developing world—a growing challenge. *N. Engl. J. Med*. 356: 213–215.

Janson, J., Soeller, W. C., Roche, P. C., Nelson, R. T., Torchia, A. J., Kreutter, D. K., and Butler, P. C. (1996) Spontaneous diabetes mellitus in transgenic mice expressing human islet amyloid polypeptide. *Proc. Natl. Acad. Sci. U.S.A*. 93, 7283–7288

Kahn. S.E, Andrikopoulos. S and Verchere, C.B. (1999) Islet amyloid: a long-recognized but underappreciated pathological feature of type 2 diabetes. *Diabetes* 48, 241–253.

Kitchen, D.B., Decornez, H., Furr, J.R. & Bajorath, J. (2007) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature reviews Drug discovery*. 3(11):935-949.

Letunic.I, Doerks.T and Bork.P (2009) SMART 6: recent updates and new developments. *Nucleic Acids Res.* 37: 229/32.

Madhu.S and Mahesh.P. (2015) Sequence analysis of Semaphorin in tumor progression: An Insilico approach *International Journal of Asian Academic Research of Multidisciplinary* .1(29): 407-423.

Mahesh.P, Divya.P, Akshatha.M, Prathima.R and Lava Kumar.C. (2015) Insilico characterization and phylogenetic analysis of novel probiotic bacteria in honey bees *International Journal of Asian Academic Research of Multidisciplinary* .1(32): 337-357.

Mandeep Singh, Mahesh Pattabhiramaiah and Swetha.R (2016) Bio-computational Analysis of fusicoccin from Tomato Plant. *International Journal of Cell Science and Biotechnology.* Volume 5; Page no 1-11.

Moriarty.D.F, and Raleigh.D.P. (1999) Effects of sequential proline substitutions on amyloid formation by human amylin 20-29. Biochemistry 38: 1811–1818.

Mugilan, A., Ajitha, M.C., Devi & Thinagar. (2010) In silico Secondary Structure Prediction Method (Kalasalingam University Structure Prediction Method) using Comparative Analysis. *Trends in Bioinformatics*. 3(1):11-19.

Nanga.R.P, Brender.J.R, Xu.J, Veglia.G, Ramamoorthy.A. (2008) Structures of rat and human islet amyloid polypeptide IAPP (1-19) in micelles by NMR spectroscopy. *Biochemistry. 47,* (48), 12689-12697.

Prashanth.V.T, Uddhav.S.C, Madura.S.M, Vishal.P.D and Renuka.R.K. (2010) Secondary Structure Prediction and Phylogenetic Analysis of salt Tolerant Proteins. *Global Journal of molecular Sciences*. 5(1):30-36.

Praveen Kumar. K.S. & Mahesh Pattabhiramaiah. (2015) Sequence analysis of basic phospholipase A2 (neurotoxin) as a potential drug target: an in silico approach. *International Journal of Engineering Research & General Science* Volume 3, Issue 3.1057-1067.

Remmert, M., Biegert, A., Hauser, A. & Soding, J. (2012) HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*, 9, 173-175.

Sali, A. & Blundell, T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*, 234, 779-815.

Schmitz.O, Brock.B and Rungby.J. (2004) Amylin agonists: A novel approach in the treatment of diabetes. *Diabetes* 53: S233–S238.

Schultz.J, Milpetz.F, Bork.P and Ponting.C.P. (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* 95 (11): 5857/64.

Tenidis.K, Waldner.M, Bernhagen.J, Fischle.W, Bergmann.M, Weber.M, Merkle.M.L, Voelter.W, Brunner.H, and Kapurniotu.A. (2000) Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties. *J. Mol. Biol*. 295: 1055–1071.

Tsetlin.V.I, and Hucho.F. (2004) Snake and snail toxins acting on nicotinic acetylcholine receptors. Fundamental aspects and sciences, *Bioinformatics*. 4: 53-62.

Vishwanath.K.V, Mahesh Pattabhiramaiah and Keerthi Ramesh (2015) Bio computational analysis of protein sequence of sickle cell anemia. *International Journal of Engineering Research and Generic Science*. Volume 1; Issue 1; Page no 63-73.

Westermark.P, Engstrom.U, Johnson.K.H, Westermark.G.T, and Betsholtz.C. (1990) Islet amyloid polypeptide: Pinpointing amino acid residues linked to amyloid fibril formation. *Proc. Natl. Acad. Sci*. 87: 5036–5040.

Yan.L.M, Tatarek-Nossol.M, Velkova.A, Kazantzis.A and Kapurniotu.A. (2006) Design of a mimic of non amyloidogenic and bioactive human islet amyloid polypeptide (IAPP) as nanomolar affinity inhibitor of IAPP cytotoxic fibrillogenesis. *Proc. Natl. Acad. Sci*. 103: 2046–2051.

Young.A, Denaro.M. Roles of amylin in diabetes and in regulation of nutrient load. Nutrition; 14: 524–527 1998.

Zanuy, B.Ma and R.Nussinov, (2003) Short peptide amyloid organization: stabilities and conformations of the islet amyloid peptide NFGAIL, *Biophys. J*. 84 1884e1894.

Zlkai, A. (1980) Thermo stability and aliphatic index of globular proteins. *J. Biochem*. 88:1895-1898.