EUROPEAN ACADEMIC RESEARCH
ISSN 2286-4822
www.euacademic.org

EUROPEAN ACADEMIC RESEARCH
Vol. II, Issue 11/ February 2015

Impact Factor: 3.1 (UIF)
DRJI Value: 5.9 (B+)

"In-Silico Design and Analysis of CYYR1- A Putative Gene/Protein Involved in Neuroendocrine Tumors."

KIRTI BHADHADHARA[1]
B. Tech Bioinformatics
Amity Institute of Biotechnology
Amity University, Rajasthan, India
NIDHI MATHUR
AMRENDRA NATH PATHAK
Amity University Rajasthan, India

Abstract:

Neuroendocrine tumorsare neoplasms that arise from cells of the endocrine (hormonal) and nervous systems. By in-depth analysis of the chromosome 21, segment 40/105 (21q21.1), a novel human gene was identified, which was previously never found to have any coding region, which is the CYYR1 gene/protein of unknown functionality.The gene consists of four exons and spans about 107 kb, including a very large intron of 85.8 kb.. The most prominent feature identified in the protein family is a central, unique cysteine and tyrosine-rich domain, which is strongly conserved from lower vertebrates (fishes) to humans but is absent in bacteria and invertebrates. In our present study, we have tried to study the protein in detail using insilico analysis and modeled the protein, along with analyzing its interactions with putative interacting proteins. It was modelled using de novo pathway as there was no structure more than 34.783% identity. The modelled structure had overall quality to of 38% which was enhanced upto 86.047 % by using loop modeling concept by ModLoop software. The Interactome was predicted using Genemania which gave us average interactions with many genes but among them most prominent was FLT1 and CALCR1. The docking studies revealed that there might be interaction between FLT1 which is a kinase enzyme, while interaction with CALCR1 was not of prominence. The interaction thus provides us an idea that there might be some kind of interaction between FLT1 and CYYR1 which has an important role in activating the gene/protein leading to the disease that it causes.

Key Words: Neuroendocrine tumor, Ab-initio-Modeling, Docking, Interaction analysis.

Introduction:

Neuroendocrine tumorsare neoplasms that arise from cells of the endocrine (hormonal) and nervous systems (Modlin IM et al., 2008). Many are benign, while some are malignant. By in-depth analysis of the chromosome 21, segment 40/105 (21q21.1), a novel human gene was identified, which was previously never found to have any coding region (Vitale L, Frabetti F et al ,

[1] Corresponding author: kirti.amity@gmail.com

Kulke MH, Siu LL, Tepper JE, Fisher G, Jaffe D, et al., 2011). The gene/protein derived from brain cells, when translated provided a 154 amino acid product which had no similarity with any known protein. The gene has been named cysteine and tyrosine-rich protein 1 gene (CYYR1).

The gene consists of four exons and spans about 107 kb, including a very large intron of 85.8 kb. Analysis of expressed sequence tags shows high CYYR1 expression in cells belonging to the amine precursor uptake and decarboxylation system (Su-Chen Li et al, 2012). Sequence and phylogenetic analysis led to identification of several genes encoding CYYR1 homologous proteins. The most prominent feature identified in the protein family is a central, unique cysteine and tyrosine-rich domain, which is strongly conserved from lower vertebrates (fishes) to humans but is absent in bacteria and invertebrates. These cells may have a link for the origin of neuroendocrine (NE) tumors (Kloppel G et al., 2004).The gene has not been studied in detail. The involvement ofCYYR1in neuroendocrine tumor though has not been experimentally proved, but various previously done studies have shown that it is somewhere involved in the disease (Donna Karolchik, et.al, 2014). In our present study, we have tried to study the protein in detail using insilico analysis and modeled the protein, along with analyzing its interactions with putative interacting proteins. The protein is modeled using de novo pathway as it had no structural similarity with any of the known proteins. Using literature survey, we got the molecular weight=16.6 kdal and the isoelectric point=8.28.

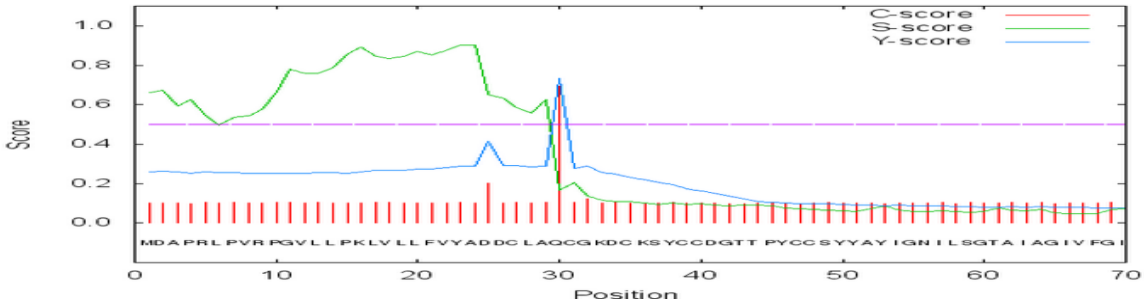
It is assumed that NE tumors can originate from normal NE cells.NE tumors are divided, according to the WHO(WorldHealthOrganization) classification, into well differentiated tumors or carcinomas, poorly differentiated carcinomas and mixed endocrine-exocrine carcinomas, according to their histological and cytological features(Pavel ME, Hainsworth JD, Baudin E, Peeters M, Horsch D, et al.,2011).

First, the CYYR1 mRNA coding sequence (CDS) was studied in a large series of NEtumors and a P111S mutation in the only neck-derived tumor was identified. Both isoforms, which encode two polypeptides–differing in terms of the absence or presence of one amino acid were found in normal and neo-plastic tissues. The protein was structured and its interaction analysis was performed. The other interacting partners were docked to study the interactions between the CYYR1 and the putatively interacting protein. No specific protein was found to interact with the CYYR1 protein; all proteins were known to be hypothetically predicted to interact with the CYYR1 protein of unknown function.

Methodology:

UniProt is a comprehensive, high-quality and freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects (Magrane M. and the UniProt consortium, 2009). It contains a large amount of information about the biological function of proteins derived from the research literature (Design and implementation of the UniProt websiteBMC Bioinformatics, 10:136, 2009). The UniProt ID-Q96J86 was taken, which was a reviewed entry in the database with 154 amino acid residues in the protein/gene.

Using SignalP tool, three domains were determined: Signal Peptide domain from position 1-29, transmembrane domain from position 62-82 and the Poly-Proline domain from position 144-149.




```
# Measure  Position  Value    Cutoff   signal peptide?
  max. C    30       0.695
  max. Y    30       0.735
  max. S    24       0.904
  mean S     1-29    0.715
       D     1-29    0.727   0.500   YES
Name=Sequence   SP='YES' Cleavage site between pos. 29 and 30: CLA-QC D=0.727 D-cutoff=0.500 Networks=SignalP-TM
```

Figure-1 The amino acids from position 1-29 are a part of signal peptides.

Sequence Analysis and Structure Prediction:

The Pfam database is a large collection of protein families, each represented by multiple sequence alignments and hidden Markov models (HMMs) (M. Punta et al, 2014).Proteins aregenerally composed of one or more functional regions, commonly termed domains. Different combinations of domains give rise to the diverse range of proteins found in nature. The domain was analyzed using Pfam 27.0, which shows that the domain is DUF26268.

Jpred is a Protein Secondary Structure Prediction server and has been in operation since approximately 1998(Cole C, Barber JD and Barton GJ. (2008). Jpred incorporates the Jnet algorithm in order to make more accurate predictions. In addition to protein secondary structure Jpred also makes predictions on Solvent Accessibility and Coiled-coil regions (Lupas method).The secondary structure analysis was done.

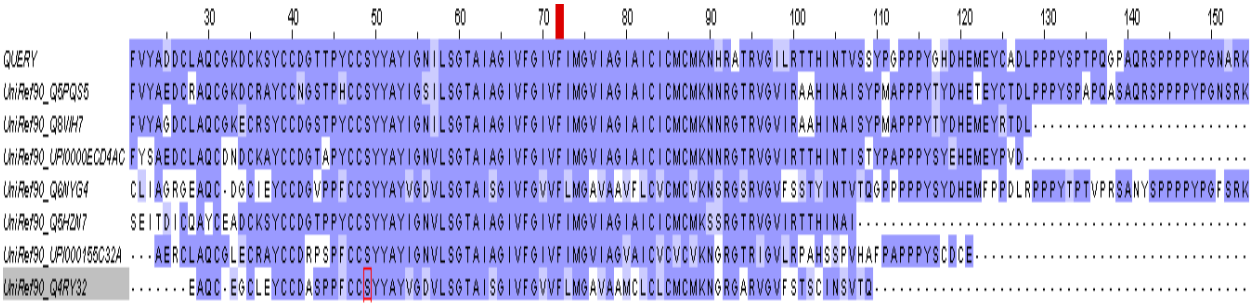


Figure-2. Secondary Structure Analysis of CYYR1 gene/protein

The Protein Data Bank (PDB) is a repository for the three-dimensional structural data of large biological molecules, such as proteins andacids (Luthy Bowie JU, Eisenberg D.1992, F.C.Bernstein et al). The data, typically obtained by X-ray crystallography or NMR spectroscopy andsubmittedbybiologists and biochemists from around the world. The structure for Cytosine and tyrosine rich protein1 was not present in the PDB entry, so the structure was modeled. BLAST is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. A BLAST search

enables a researcher to compare a query sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold (Altschul, S.F. et al, 1990).

Protein BLAST was done with the database selected as PDB-database, which shows the maximum identity as 39% and 37%, with the PDB-ID as 4BZJ_A and 2LFC_A respectively as shown in table 1. As the identity was less than 40% Schrodinger-PRIME rejected the homology modeling of the protein (Biologics Suite 2014-2: BioLuminate, version 1.5, Schrödinger, LLC, New York, NY, 2014.). Hence Ab-initio modeling using-Phyre2 server was done to predict its structure (Protein structure prediction on the web, 2009).

Table-1: Protein Blast results which were used as template for structure modeling.

Description	Max-Score	Total Score	Query Coverage	E-Value	Identity	Accession
Chain A, The Structure of Copii Coated Assembled on Membrane (Saccharomyces Cerevisiae)	28.5	28.5	24%	2.8	39%	4BZJA
Chain A, A solution NMR str. of Fumarate reductase flavoprotein subunit from lactobacillus plantarum	26.9	26.9	19%	6.0	37%	2LFCA

After prediction the model was optimized by energy minimization using Swiss–Pdb Viewer.Swiss–Pdb Viewer (Spdbv) is an application that provides a user friendly interface allowing analyzing several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts (Kiefer F et al-2009,). The energy minimization was done using the Spdbv software, until the energy reached a constant energy value.The structure was further optimized and modeled using, automated modeling of loops by Mod Loop (A. Fiser, and A. Sali, Bioinformatics, 2003).The predicted structure as shown in figure 3, has a big alpha helix with two small beta sheets and loops clearly visible.

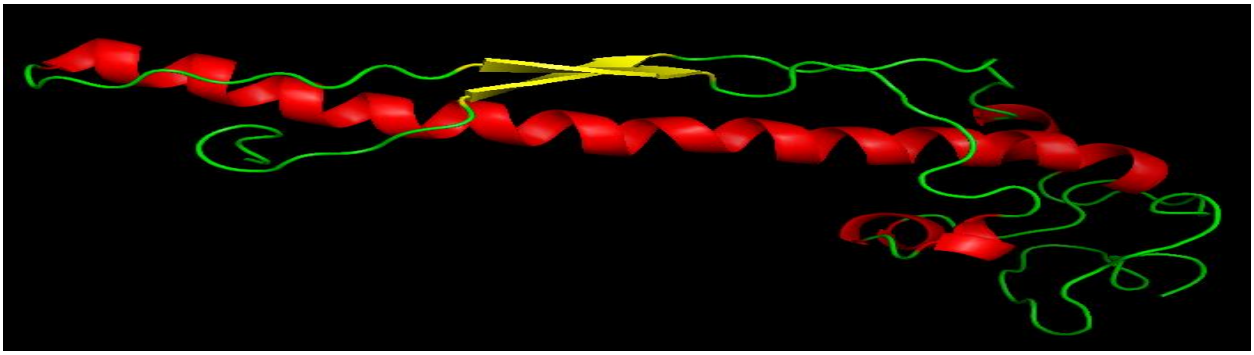


Figure-3_Representation of the modeled structure using (The PYMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger)

Interaction Analysis: -Gene Mania finds other genes that are related to a set of input genes, using a very large set of functional association data. Association data include protein and genetic interactions, pathways, co-expression, co-localization and protein domain similarity. (David Warde-Farley, et.al, 2010). The CYYR1 gene/protein interaction analysis search was made using

Gene Mania. The table 2 depicts all the interactions predicted by Gene Mania for CYYR1, which are diagrammatically represented in figure 4, where the interactions are connected using the flow diagram.

Table-2 Interaction analysis of CYYR1 gene/protein

	Networks	**Genes/Proteins involved**	**Percent Score**
1.	Co-expression	MCOLN3,GCA,BMPER,CD93,CLEC14A,EMCN,FLT1,GJA4,LDB2,CALCRL,HSA,CLEC14A,MLOLN3,HSPA12B,C20orf160,ERG,MYCT1,APBB2, L3MBTL3,FGD5.	35.49%
2.	Pathway	FLT1,ERG	10.41%
3.	Co-localization	MCOLN3,CNPPD1,PHF10,FBX018,EMCN,FLT1,LDB2,ERG,MYCT1,CALCRL	6.37%
4.	Shared Protein Domains	CD93,CLECMA	1.54%
5.	Genetic Interactions	MCOLN3,GCA,PHF10,FBX018,EMCN,FLT1,CD93,GCA,HSPA12B,FGD5,ERG,MYCT1,APBB2	1.51%

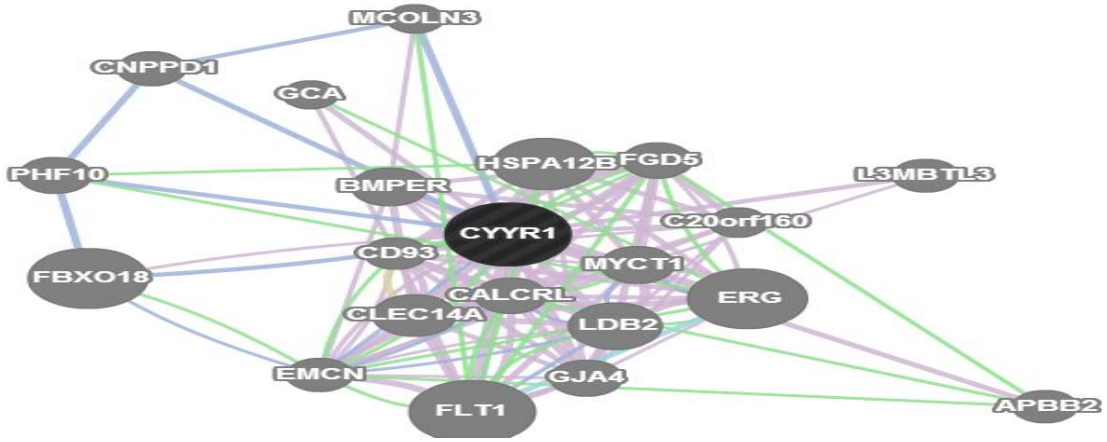


Figure-4 Interaction Map of CYYR1 gene/protein, the purple line represents co-expression, green line represents pathway and the blue line represents co-localization.

The physical interactions are not known yet for the protein , only co-localized and pathway based interactions are known , which are also putative.

Docking Studies

The Active Sites were Predicted and Analyzed using Dog site scorer for CYYR1 to recognize the cavities were another molecule should be interacting with the protein. (Center for Bioinformatics-Universitat of Hamburg, 2003). The top six cavities as identified by the tool are specified in the table 3.

Table-3 Active Sites Prediction of CYYR1 gene/protein

Name	Volume	Surface	Lipo surface	Depth	Simple Score
P0	2199.17	3428.85	2602.77	30.77	0.72
P1	250.94	689.05	441.90	8.28	0.14
P2	227.97	496.94	375.39	13.68	0.10
P3	223.62	519.63	428.20	10.20	0.13
P4	155.84	368.30	298.68	9.28	0.04
P5	131.20	385.23	316.36	8.11	0.00

Hex is an interactive protein docking and molecular superimposition program, written by Dave Ritchie (Ota, Suzuki Y. et al, 2013). Hex understands protein and DNA structures in PDB format, and it can also read small-molecule SDF files.

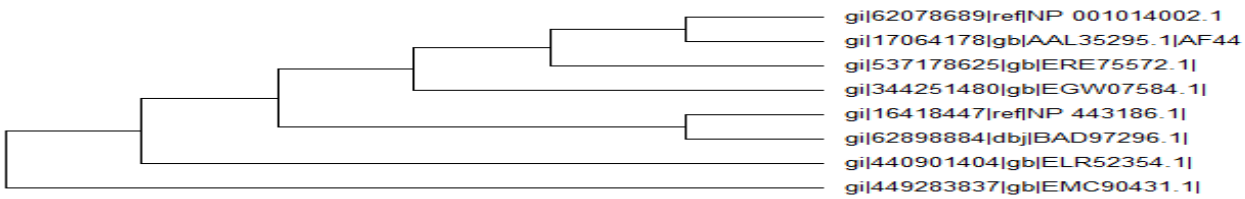
The maximum probability of interaction of CYYR1 was predicted to be with CALCR1 (3AQF.PDB) and FLT1 (1FLT.PDB). The 3AQF.PDB is a Calcitonin receptor and 1FLT.PDB is a Tyrosine-protein kinase enzyme. The protein-protein interaction was studied using the Hex docking software. The CYYR1.PDB (modeled by Phyre2 server) was docked with 3AQF.PDB and 1FLT.PDB , whose complexes were visualized in PYMOL to study the interaction.

Phylogenetic Analysis

MEGA is an integrated tool for conducting sequence alignment, inferring phylogenetic trees, estimating divergence times, mining online databases, estimating rates of molecular evolution, inferring ancestral sequences, and testing evolutionary hypotheses (Tamura K et al, 2013). MEGA is used by biologists in a large number of laboratories for reconstructing the evolutionary histories of species and inferring the extent and nature of the selective forces shaping the evolution of genes and species.

	1	2	3	4	5	6	7	8
1. gi|440901404|gb|ELR52354.1|								
2. gi|537178625|gb|ERE75572.1|	0.045							
3. gi|344251480|gb|EGW07584.1|	0.034	0.011						
4. gi|449283837|gb|EMC90431.1|	0.129	0.167	0.154					
5. gi|16418447|ref|NP 443186.1|	0.104	0.141	0.129	0.234				
6. gi|62898884|dbj|BAD97296.1|	0.104	0.141	0.129	0.234	0.000			
7. gi|62078689|ref|NP 001014002.1	0.092	0.045	0.057	0.220	0.180	0.180		
8. gi|17064178|gb|AAL35295.1|AF44	0.116	0.068	0.080	0.234	0.193	0.193	0.045	

Figure-5(a) Pairwise Distances



(b)Bootstrap Phylogenetic Tree of CYYR1 gene/protein

Tajima's test of neutrality (Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism, 1989) which compares the number of segregating sites per site with the nucleotide diversity was conducted on CYYR1. The analysis involved 8 amino acid sequences. All positions

containing gaps and missing data were eliminated. There were a total of 91 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.

Table. Results from Tajima's Neutrality Test [1]

m	*S*	p_s	Θ	π	*D*
8	26	0.285714	0.110193	0.114207	0.192031

Figure-6:_ TaximaTest of Neutrality

Abbreviations: m = number of sequences, n = total number of sites, S = Number of segregating sites, ps = S/n, T = ps/a1, p = nucleotide diversity, and D is the Tajima test statistic. (Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0).

The CYYR1 as predicted to be interacting with the Tyrosine Kinase (FLT1), the protein was put through phosphorylation testing. The NetPhos tool was used to identify the probable phosphorylating sites.

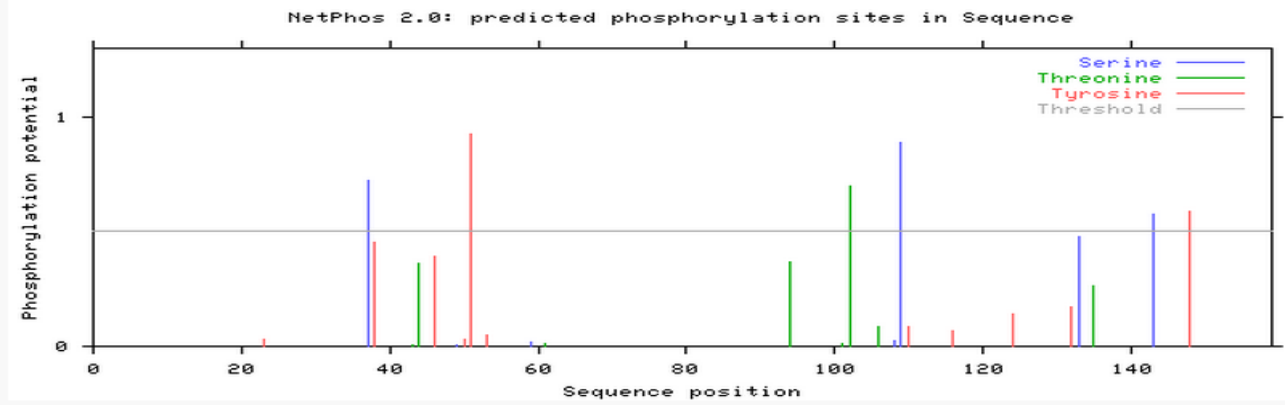


Figure-7:_Predicted Phosphorylation sites

As evident from the literature that there is mutation at the 111th position of CYYR1 protein in which Proline residue is replaced by a Serine residue. The effect of the mutation was studied on the structure by redesigning the structure of mutated CYYR1 protein using Phyre-2 server again. The predicted structure was then visualized in PYMOL as shown in the figure 11.

Validation:-VADAR is a comprehensive web server for quantitative protein structure evaluation. It accepts Protein Data Bank (PDB) formatted files or PDB accession numbers as input and calculates, identifies, graphs, reports and/or evaluates a large number (>30) of key structural parameters both for individual residues and for the entire protein (Leigh Willard et al., 2003). These include excluded volume, accessible surface area, backbone and side chain dihedral angles, secondary structure, hydrogen bonding partners, hydrogen bond energies, steric quality, solvation free energy as well as local and overall fold quality. The modeled protein was evaluated on for the Ramachandran Plot and overall quality of the structure.

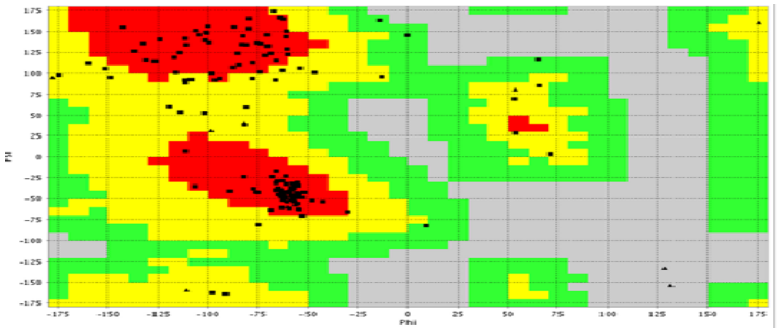


Figure-8: Ramachandran Plot

Number of residues in favored region: 130 (85.5%)
Number of residues in allowed region: 12 (7.9%)
Number of residues in outlier region: 10 (6.6%)

Statistic	Observed	Expected
# Helix	52 (33%)	-
# Beta	25 (16%)	-
# Coil	77 (50%)	-
# Turn	28 (18%)	-

Figure-9: Statistics of the secondary structure elements present in CYYR1 as predicted by Ab initio method using Phyre2.

Result and Discussion:

CYYR1 an unknown protein/gene definitely has some or the other connection to the neuroendocrine tumor. Though experimentally it is not yet verified, but the computational analysis thus performed provides with the numerous important findings. Three domains were analyzed namely as Signal peptide domain (1-29), Transmembrane domain (62-82), Poly Proline domain (144-149).The structure was predicted using Phyre2 server, as it showed very low amount of homology of the order 35% only. The predicted structure had an ERRAT score of 34.783 and finally its loop modeling enhanced the overall quality of the structure with over all enhancing the quality of the structure to 86.047%.

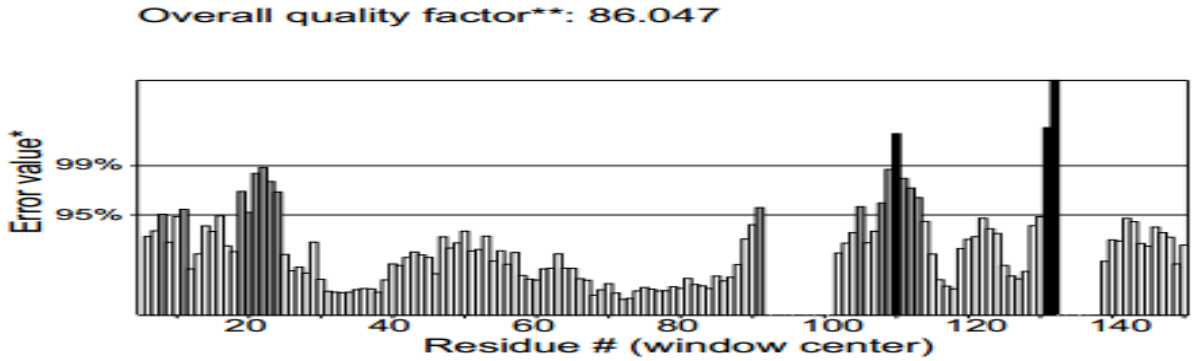


Figure-10:_ERRAT plot after energy minimization and loop modeling with overall quality factor-> 86.047

The structure contained one alpha helix and two beta sheet as shown in figure 3.

The mutations cited in literature were replacement of Proline at 111th position by Serine. When the structure was designed and analyzed it was found that due to a single mutation the two stable beta sheets are all lost which destabilizes the structure and prevents it from attaining its stable conformation, which might also effect its interaction with other interacting partners. The comparison of the structure is visualized in figure 11, where the yellow colored beta sheets are lost in the single mutant form of CYYR1 protein.

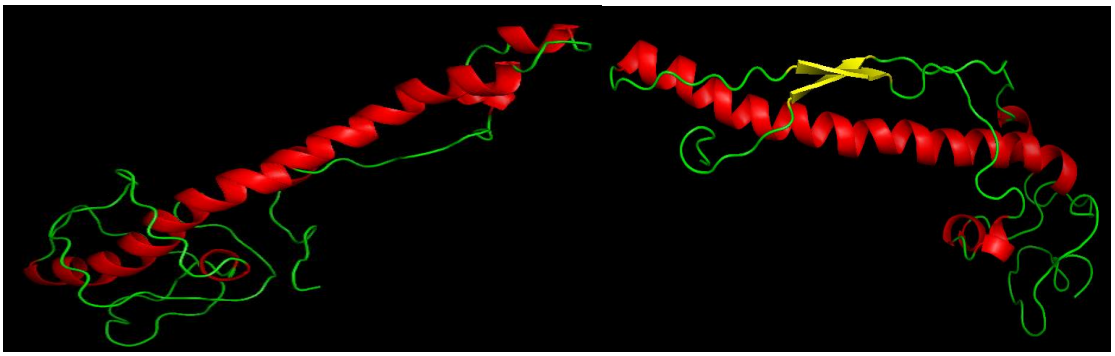


Figure-11:_P replaced by S at 111sites in sequence position (mutational study)

The entry in the UniprotKB describes 2 isoforms produced by alternative splicing. In natural variant, at 95th position Arginine replaced by Histidine, at 111th position Proline replaced by Serine.

The interaction of CYYR1 gene/protein was analyzed using Gene Mania, showing interaction with the FLT1 gene and CALCRL. The detailed study of amino acids involved in active site revealed a Volume = 2199.17, Surface = 3428.85 and Simple Score = 0.72, with the amino acids like Proline, Argenine, Leucine , Isoleucine and Methionine were found to be involved in the active site.

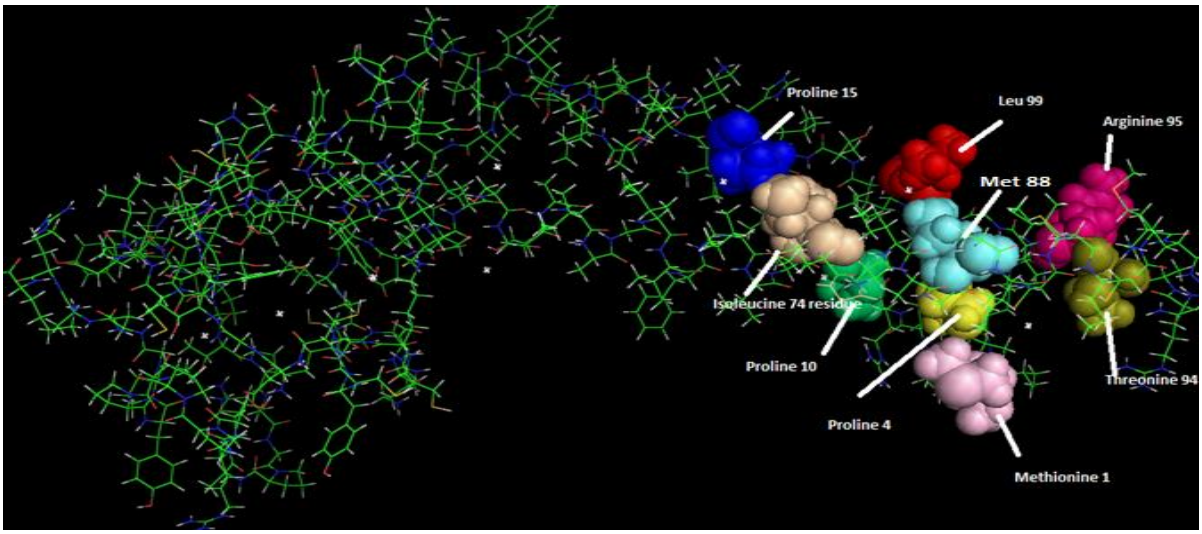


Figure-12: Active Sites Prediction with *Active Site Predictor* (B.Jayaramet al, 2011)

The protein-protein interaction was studied using docking of CYYR1 with FLT1and CALCRL using Hex. The figure 13 and 14 represents the interaction between the hypothetically predicted interacting partners, as predicted from Gene Mania. The interaction when studied in detailed

revealed that interaction between CYYR1 and FLT1 is possible and is still stable as compared to that with the CALCRL. The FLT1 is a kinase enzyme which as predicted might have some specific interaction leading to activation of this CYYR1.

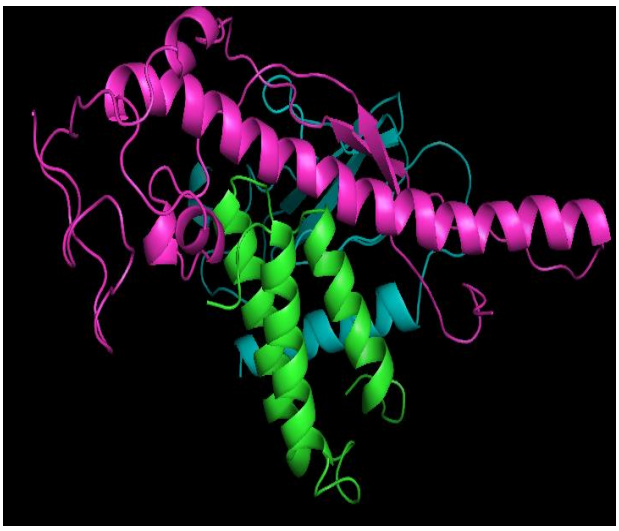
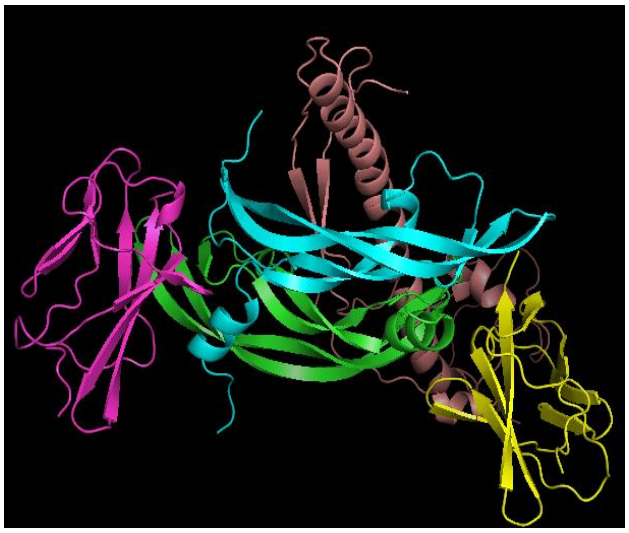


Figure-13 (a) Hex docking results with 3AQF, Evalue-> -819.90



(b) Hex docking results with FLT1, Evalue -> -910.52

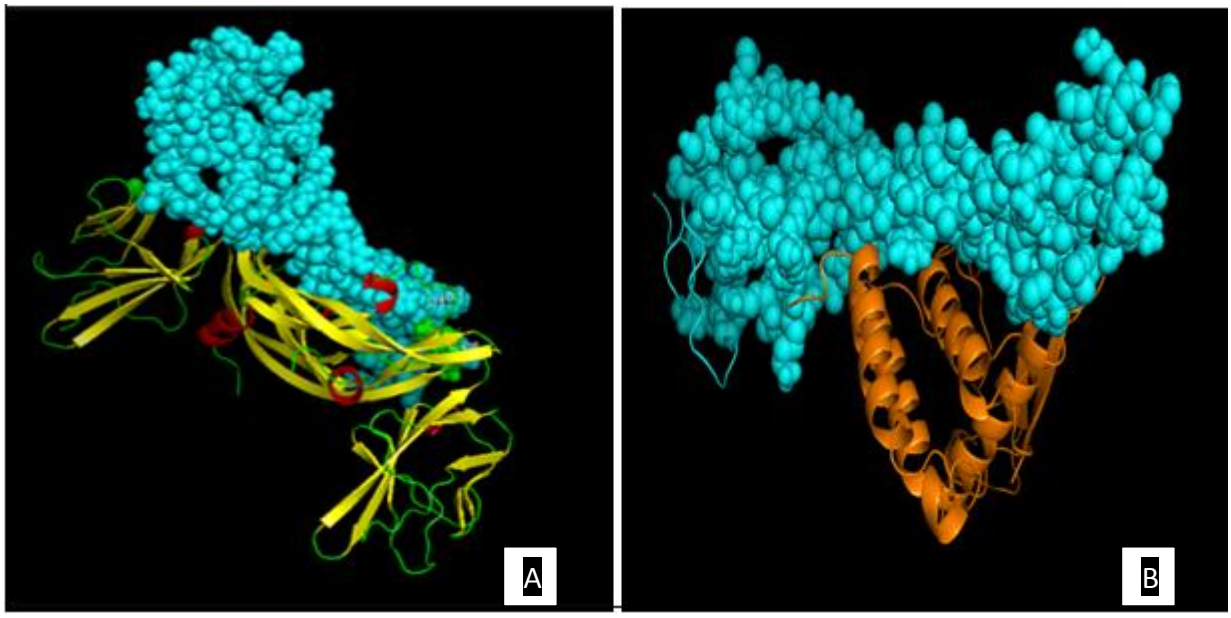


Figure 14: The snapshots from Pymol of CYYR1 interacting with FLT1 in A and with CALCRL in B.

The figure 13 and 14 clearly shows that FLT1 has very clear interaction with the CYYR1 and it tries to interact with it from three sides. The stability of the complex is also good with a energy score of -910.52 KJ/mol. The interacting amino acids are Alanine, Proline , Lysine, Isoleucine, Histidine.

Mega-6 software was used for performing Pairwise Alignment between 8 co-related (phylogenetically related) sequences. TaximaTest was performed to check the mutation probability, which came as 0.192031.

The phosphorylated sites (Serine, Threonine, Tyrosine and Threshold) were analyzed using NetPhos, as shown in figure 7 further clarifies that there could be a possible interaction between FLT1 phosphorylating the CYYR1 making it active to participate in some important pathway involved in neuro-endocrine tumor or vice versa. The mutated The Proline was replaced by Serine at 111^{th} position, it was observed that the 2 beta sheets were replaced by the alpha helix, which makes it a destabilized structure which could lead to its inactivation and thus hindering the activity of CYYR1.

The Ramachandran plot also clarifies that it has more of a coiled-coiled structure having 50% of it making it a transmembrane protein which when activated performs some or the other task in trans-membrane related transferring of some important molecule involved in the neuro-endocrine tumor. The structure though is not completely stable as specified by the plot with around 10 amino acids which are the present in the outlier area. The structure needs more improvement to be studied in detail to get to a conclusion related to its involvement in some important pathway leading to neuro-endocrine tumor.

The area needs to be further studied in detail with experimentally designing the CYYR1 structure whose interaction can be further studied in detail to come to a final conclusion leading to interpretation of the exact role that it plays in neuro-endocrine tumors. The finding of involvement of CYYR1 in this disease is new which should be pondered upon to target the specific molecule involved in this disease, thus providing a new insight into the understanding of the various cascades of steps leading to neuro-endocrine tumor.

REFERENCES:

1. Su-Chen Li, Cecile Martijn, TaoCui, Ahmed Essaghir, Rau, M. Luque, Jean-Baptiste DE Moulin, Justo P. Castan, Kjell Oberg, Valeria, Giandomenico "The Somatostatin Analogue Octreotide Inhibits Growth of Small Intestine Neuroendocrine Tumor Cells".
2. Vitale L, Frabetti F, Huntsman SA, Canaider S, Casadei R, Lenzi L, Facchin F, Carinci P, Zannotti M, Coppola D, Strippoli Sequence, "subtle" alternative splicing and expression of the CYYR1 (cysteine/tyrosine-rich 1) mRNA in human neuroendocrine tumors".
3. Vitale L, Casadei R, Canaider S, Lenzi L, Strippoli P, D'Addabbo P, Giannone S, Carinci P, Zannotti M "Cysteine and tyrosine-rich 1 (CYYR1), a novel unpredicted gene on human chromosome 21, encodes a cysteine and tyrosine-rich protein and defines a new family of highly conserved vertebrate-specific genes".
4. Kloppel G, Perren A, Heitz PU (2004) The gastroenteropancreatic neuroendocrine cell system and its tumors: the WHO classification. Ann N Y Acad Sci 1014: 13–2.
5. Modlin IM, Oberg K, Chung DC, Jensen RT, de Herder WW, et al. (2008) Gastroenteropancreatic neuroendocrine tumors. Lancet Oncol 9: 61–72
6. Kulke MH, Siu LL, Tepper JE, Fisher G, Jaffe D, et al. (2011) Future Directions in the Treatment of Neuroendocrine Tumors: Consensus Report of the National Cancer Institute Neuroendocrine Tumor Clinical Trials Planning Meeting. J Clin Oncol 29: 934–943
7. Pavel ME, Hainsworth JD, Baudin E, Peeters M, Horsch D, et al. (2011) Everolimus plus octreotide long-acting repeatable for the treatment of advanced neuroendocrine tumors

associated with carcinoid syndrome (RADIANT-2): a randomized, placebo-controlled, phase 3 study. Lancet 378: 2005–2012

8. Strosberg J, Kvols L (2010) Ant proliferative effect of somatostatin analogs in gastroenteropancreatic neuroendocrine tumors. World J Gastroenterol 16: 2963–2970.
9. Oberg KE, Reubi JC, Kwekkeboom DJ, Krenning EP (2010) Role of somatostatins in gastroenteropancreatic neuroendocrine tumor development and therapyGastroenterology 139: 742–753, 753 e741.
10. Culler MD, Oberg K, Arnold R, Krenning EP, Sevilla I, et al. (2011) Somatostatin analogs for the treatment of neuroendocrine tumors. Cancer Metastasis Rev 30: 9–
11. Oberg KE (2011) The Management of Neuroendocrine Tumors: Current and Future Medical Therapy Options. Clin Oncol (R Coll Radiol).
12. M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Son hammer, S.R. Eddy, A. Bateman, R.D. Finn-PFam.
13. Magrane M. and the UniProt consortium UniProt Knowledgebase: a hub of integrated protein dataDatabase, 2011: bar009 (2011).
14. Infrastructure for the life sciences: design and implementation of the UniProt websiteBMC Bioinformatics, 10:136 (2009).
15. The Pfam protein families database: M. Punta, P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, R.D. FinnNucleic Acids Research (2014)
16. Tamura K, Stecher Peterson D, Filipski, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Molecular Biology and Evolution 30:2725-2729.
17. F.C.Bernstein, T.F.Koetzle, G.J.Williams, E.E.Meyer Jr., M.D.Brice, J.R.Rodgers, O.Kennard, T.Shimanouchi, M.Tasumi, "The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures," J. of. Mol. Biol., 112 (1977): 535.
18. The Gene MANIA prediction server: biological network integration for gene prioritization and predicting gene function. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT.
19. Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q.
20. Nineteen additional unpredicted transcripts from human chromosome 21."
21. Reymond A., Camargo A.A., Deutsch S., Stevenson B.J., Parmigiani R.B., Ucla C., Bettoni F., Rossier C., Lyle R., Guipponi M., de Souza S., Iseli C., Jongeneel C.V., Bucher P., Simpson A.J.G., Antonarakis S.E.Genomics 79:824-832(2002).
22. Complete sequencing and characterization of 21,243 full-length human cDNAs. Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., ObayashiM., Nishi T., Shibahara T., Tanaka T., Ishii S. Sugano S.Nat. Genet. 36:40-45(2004).
23. The DNA sequence of human chromosome 21."Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,Park H.-S.,Toyoda A.,Ishii K.,Totoki Y.,Choi D.-K.,Groner Y.,Soeda E.,Ohki M.,Takagi T.,Sakaki Y.,Taudien S.,Blechschmidt K.,Polley A.Yaspo M.L.
24. Sequence, 'subtle' alternative splicing and expression of the CYYR1 (cysteine/tyrosine-rich 1) mRNA in human neuroendocrine tumors."Vitale L., Frabetti F., Huntsman S.A.,

Canaider S., Casadei R., Lenzi L., Facchin F., CarinciP., Zannotti M., Coppola D., Strippoli P.BMC Cancer 7:66-66(2007).

25. Laskoswki RA, MacArthur MW, Moss DS & Thorton JM. (1993). PROCHECK: a program to check the stereo chemical quality of protein structures. J. Appl. Cryst. 26, 283-291.
26. Hooft, RWW, Vriend G, Sander C, Abola EE. (1996).Errors in protein structures. 381,272-272.
27. Colovos C, Yeates TO. (1993). Verification of protein structures: patterns of non-bonded atomic interactions. Protein Sci. 2, 1511-1519.
28. Luthy Bowie JU, Eisenberg D. (1992).Assessment of protein models with three-dimensional profiles. Nature 356, 83-85.
29. Ponitus J, Richelle J, Wodak SJ. (1996).Deviations from standard atomic volumes as a quality measure for protein crystal structures. J.Mol. Biol. 264,121-136.
30. Vaguine AA, Richelle J, Wodak SJ. (1999).SFCHECK: a unified set of procedure for evaluating the quality of macromolecular structure-factor data and their agreement with atomic model. Acta Cryst. D55, 191-205.
31. The Jpred 3 secondary structure prediction server-Christin, Cole, JonathanD. Barber and Geoffrey J. Barton
32. VADAR: a web server for quantitative evaluation of protein structure quality. Leigh Willard, Anuj Ranjan1, Haiyan Zhang, Hassan Monzavi1, Robert F. Boyko, Brian D. Sykes and David S. Wishart.
33. Protein structure prediction on the web: a case study using the Phyre server. Kelle LA and Sternberg MJE. Nature Protocols 4,363-371(2009).
34. Schomburg, K.T.; Bietz, S.; Briem, H.; Henzler, A.M.; Urbaczek, S.; Rarey, M. (2014). Facing the Challenges of Structure-Based Target Prediction by Inverse Virtual Screening. Journal of Chemical Information and Modeling.
35. Protein Docking Using Case-Based Reasoning. A.W. Ghoorah, M. Smail-Tabbone, M.-D. Devignes, D.W. Ritchie, (2013). Proteins: Structure, Function, Bioinformatics, DOI:10.1002/prot.24433.
36. Larkin MA, Black shields G, Brown NP, Chenna R, McGettigan PA, McWilliams H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ and Higgins DG Bioinformatics 2007 23(21): 2947-2948. doi:10.1093/bioinformatics/btm404.
37. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T (2009).The SWISS-MODEL Repository and associated resources. Nucleic Acids Res. 37, D387-D392.
38. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410. PubMed
39. The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
40. ö Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A., "A Hierarchical Approach to All-Atom Protein Loop Prediction," Proteins: Structure, Function and Bioinformatics, 2004, 55, 351-367
41. Ö Bell, J.A., Cao, Y., Gunn, J.R., Day, T., Gallicchio, E., Zhou, Z., Levy, R. and Farid, R., "PrimeX and the Schrödinger Computational Chemistry Suite of Programs," International Tables for Crystallography, 2012, F (18), 534-538
42. Tajima F. (1989). Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

43. Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution30: 2725-2729.
44. Dog Site Scorer, Center for Bioinformatics-Universitat 0f Hamburg, 2003.
45. Donna Karolchik,Galt P. Barber, Jonathan Casper, Hiram Clawson, Melissa S. Cline, Mark Die khans, Timothy R. Dreszer1,Pauline A. Fujita1,Luvina Guruvadoo1,Maximilian Haeussler1,Rachel A. Harte1,Steve Heitner1,Angie S. Hinrichs1,Katrina Learned1,Brian T. Lee1,Chin H. Li1,Brian J. Raney1,Brooke Rhead2,Kate R. Rosenbloom1,Cricket A. Sloan3,Matthew L. Speir1,Ann S. Zweig1, David Haussler1,4,Robert M. Kuhn1 and W. James Kent1.The UCSC Genome Browser database: 2014 update, Nucleic Acid Research, 2014.
46. Biologics Suite 2014-2: BioLuminate, version 1.5, Schrödinger, LLC, New York, NY, 2014.
47. Tanya Singh, D. Biswas, B.Jayaram, AADS - An automated active site identification, docking and scoring protocol for protein targets based on physico-chemical descriptors., 2011, J. Chem. Inf. Modeling,51 (10), 2515-2527,DOI: 10.1021/ci200193z.