IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

STUDYING AND COMPARING THE PHYSICO CHEMICAL PROPERTIES OF IRGM PROTEIN IN DIFFERENT ORGANISMS AND INTERPRETING ITS SIGNIFICANCE

Writing Error Free code to develop database driven application and Creating a predictor to study the advanced properties of the protein in retrieving numerical and script data

[1]Divya Peddapalyam,[2]Reetika Mishra,[3]Dr.Satya Narayana Chowdary,[4]Dr.Madhu Malleshappa

[1]Msc.Biotechnology,[2]Msc.Biotechnology,[3]M.Tech;PhD-Computer science,[4]Msc, PhD,Post-Doc

[1]Department of Biotechnology, Garden City University, Bengaluru,Karnataka,India

[2]Department of Biotechnology,Garden City University,Bengaluru,Karnataka,India

[3]Department of Computer science,Hyderabad,Telangana,India

[4]Department of Biotechnology,Garden City University,Bengaluru,Karnataka,India

Abstract: This examination has been attempted to explore the physical and chemical properties of the IRGM protein in various organisms.The capacity of the protein is firmly related with its subcellular location.With the achievement of the human genome project and fast increment in number of recently discovered protein sequences going into the information banks,It is profoundly attractive to build up a robotized technique for foreseeing subcellular area of the proteins.The foundation of such an indicator will no uncertainty facilitate the usefulness assurance of recently discovered proteins and procedure of organizing qualities and proteins recognized by genomic endeavors as potential atomic focuses for drug design.Based on the idea of the pseudo aminoacid and by studying the properties of the proteins of the pseudo genes,the digital signal processing approach is used to incorporate sequence order effect.One of the wonderful merits in doing so is that many existing instruments in arithmetic and designing can direct utilized in the protein sub cell location.The results consequently acquired are very encouraging.Its is foreseen that the advanced sign preparing may fill in as helpful vehicle for some other protein science zones as well.Our study plans to explore broadly the machine learning based methods to anticipate protein properties to offer a general just as point by point understanding for specialists in the field.Some of the models present adequate expectation exhibitions and advantageous client interfaces.These models can be considered as important devices to anticipate protein properties results before performing genuine research center experiments,thus saving labour,time and cost. In the table named"REQUIRED SOFTWARE PACKAGES"we are providing all required software packages for studying the protein properties in an advanced manner.But for this study we have used limited softwares for studying the advanced protein properties of IRGM gene and remaining software approaches we have used as an instance for creating protein prediction tool to study advanced protein properties.Along with this we also written a code for data base driven application to storethedata ,manage,retrieve,update in future.

Index Terms **– IRGM Protein,Functionality,Digital signaling,Data base driven application,Sub cellular location.**

I.INTRODUCTION:

The IRGM quality gives directions to making a protein that assumes a significant job in the insusceptible framework. This protein is engaged with a procedure called autophagy, which cells use to encompass and demolish outside trespassers, for example, bacteria and viruses. In particular, the IRGM protein helps trigger autophagy in cells tainted with specific sorts of microorganisms, including the kind of microscopic organisms that causes tuberculosis. Notwithstanding shielding cells from

disease, autophagy is utilized to reuse destroyed cell parts and stall certain proteins when they are not, at this point required. This procedure likewise assumes a significant job in controlled cell passing (apoptosis).Several varieties in or close to the IRGM quality have been related with an expanded danger of creating Crohn malady. This expanded hazard has been found fundamentally in individuals of northern European heritage. IRGM

varieties change single DNA building blocks (nucleotides) in regions of DNA that may direct when and how the IRGM protein is delivered. Specialists presume that changes including the IRGM protein may disturb the autophagy procedure, keeping the resistant framework from decimating destructive microorganisms successfully. A strange insusceptible reaction to microbes in the intestinal dividers may prompt interminable aggravation and the stomach related issues normal for Crohn illness. Immunity related GTPases(IRG proteins)are one of the most grounded early opposition frameworks against intracellular pathogens.The IRG quality family contains 21 duplicates masterminded as couple quality groups on the two chromosomes in the C57BL/6 mouse genome however has been decreased to just two duplicates in people i.e;IRGC and IRGM.IRGC isn't associated with immunity,but the human IRGM quality assumes a job in autophagy focused on pulverization of the mycobacterium tuberculosis and salmonella typhimurium.Variant IRGM haplotypes have been related with expanded hazard for crohn's illness and connected with the differential articulation of the IRGM transcripts. Insusceptibility related GTPases (IRG) assume a significant job in guard against intracellular pathogens. One individual from this quality family in people, IRGM, has been as of late ensnared as a hazard factor for Crohn's malady. Nitty gritty structure of this quality family among primates and demonstrated that a large portion of the IRG quality group was erased right off the bat in primate advancement, after the disparity of the humanoids from prosimians (around 50 million years back). Similar succession examination of New World and Old World monkey species shows that the single-duplicate IRGM quality became pseudogenized because of an Alu retrotransposition occasion in the humanoid normal precursor that disturbed the open understanding casing (ORF). We find that the ORF was restored as a piece of a polymorphic stop codon in the normal progenitor of people and extraordinary chimps. Articulation examination proposes that this change happened related to the addition of an endogenous retrovirus, which adjusted the translation commencement, grafting, and articulation profile of IRGM. These information contend that the quality became pseudogenized and was then restored through a progression of complex auxiliary occasions and recommend amazing useful versatility where alleles experience different transformative weights after some time. Such dynamism in structure and advancement might be basic for a quality family secured a weapons contest with an ever-changing collection of intracellular parasites. We used a bioinformatics asset entrance worked by the SIB Swiss Institute of Bioinformatics and specifically the SIB Web Team. It is an extensible and integrative entry getting to numerous logical assets, databases and programming instruments in various everyday issues sciences. Researchers can get to a wide scope of assets in a wide range of areas, for example, proteomics, genomics, phylogenetics/advancement, life science, population genetics, and transcriptomics. The individual assets (databases, electronic and downloadable programming devices) are facilitated in a decentralized path by various gatherings of the SIB Swiss Institute of Bioinformatics and accomplice organizations. In particular, a solitary online interface gives a typical passage point to a wide scope of assets created and worked by a wide range of SIB gatherings and outside organizations. The entryway includes an inquiry work across chosen assets. Inside, the accessibility and use of assets are checked. The entrance is focused on both master clients and for individuals who are curious about a particular area in life sciences: specifically, the new web interface gives visual direction to newcomers to this asset.

II. REQUIRED SOFTWARE PACKAGES:

S.NO	DESCRIPTION	SOFTWARE PACKAGES
01	Solvent affordability /accessibility of the protein	SCRATCH-1D-release(LINUX VERSION,6.3GB)
02	One-dimensional,bidirectional,recurrent neural networks	1D-BRNN;3.3-release,(ID- BRNN_3.3.tar.gz,lin ux version,17KB)
03	Protein-Evolutionary information/sequence profiles	PROFILpro;1,2-release,(PROFILpro_ 1.2.tar.gz,LINUX VERSION,6.1GB)
04	Homology based secondary structure and solvent accessibility prediction	HOMOLpro;1.2release,(HOMOLpro_1.2.tar.gz,LI NUX VERSION,271MB)
05	Software package to evaluate the accuracy of a profile based predictor as a function of the max cosine similarity between training and test profiles	EVALpro- 1.0(44MB,LINUX/Mc Version)
06	Protein Domain Prediction	DOMpro 1.0(LINUX-VERSION)
07	Prediction of beta residue pairs,beta-strand pairs,strand alignment,Pairing-direction and beta-sheet topology	BETApro- 1.0(LINUX-VERSION)
08	Prediction of protein solubility upon over expression	SOLpro(12MB,LINUX VERSION)
09	A Novel machine learning approach for the fast and accurate prediction of side chain conformations	SIDEpro(Installation sizes-165MB;and 210MB)
10	Prediction of capsid and tail proteins	VIRAlpro;Release-1.0(LINUX VERSION,25MB)
11	Prediction of Disordered residues from the protein sequences	DISpro1.0(LINUX-VERSION)
12	Protein Model Scoring/Selection and side chain prediction.	SELECTpro- 1.0(LINUX VERSION);Installation-size 680MB,1.6GB
13	Server for Protein identification and analysis(Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A;)	-

III. RESULTS: PART-I:

	MW of amino acid	pI	%	Negatively and Positively charged residues	Absorption of light in the medium (E)and absorbance assuming cystine residues(A)	Absorption of light in the medium(E) and absorbance with-out cystine residues(A)	Unreliabiliy of the protein(I) and Relative volume occupied by aliphatic side Chains(A)	Cystiene at following positions form Disulphide bonds	Grand Avg. of the Hydro-phaticity	Solvent accessibility E-Exposed residues; B-Buried residues	Non Transmembr ane protein(α helical and β barrel)	Ordered residues (R)and probability (P)	Disordered residue (R) and probability (P)	Antigenicity (A) And solubility upon-over expression (S)
A	19810.6	5.3	99.9	- = 20 + = 16	E=1857 A=0.93	E=184 A=0.9	I=30.82 A=81.0	72,176	-0.190	E=6 B=82	α=0.00025278 β=0.0002563	R=11 P=12 .	R=07 P=4.1	A=0.87 4 S=0.518
B	46337.78	8.3	100	- = 45 + = 49	E=360 3 A=0.77	E=35 410 A=0.7 64	I=51.53 A=96.2	08,41,259, 260, 366,373,375, 377	-0.134	E =195 B = 216	α=0.04816238 β=0.0015485	R=384 P=18.48	R=27 P=22	A=0.656 S=0.759
C	44887.87	8.6	99.9	- = 46 +=51	E=41870 A=0.93	E=4137 0 A=0.922	I=49.7 A=94.1	241,242, 340, 355, 358,359	-0.172	E=17 9 B=214	α =0.3304256 β=0.0032771	R=382 P=23.928	R=11 P=9.1 8	A=0.29740 0 S=0.773055
D	14829.92	6.2	91.0	- = 11 + = 10	E=2246 0 A=1.515		I=23.57 A=72.0	.	-0.084	E=70 B=64	α =0.0133477 β =0.00400645	R=125 P=10.21	R=09 P=6.85	A=0.918725 S=0.705538
E	32838.29	8.9	99.0	- = 28 += 32	E=40575 A=1.236	E=40450 A=1.232	I= 40.34 A=82.39	175,279	-0.261	E=175 B=122	α=0.00874484 β=0.01192735	R=286 P=34.84	R=11 P=8.25	A=0.886793 S=0.590485
F	46238.37	9.07	99.9	- = 47 + =56	E=4840 0 A=1.047	E=4790 0 A=1.036	I=46.77 A=96.7	259,270, 347, 358, 374,411	-0.152	E=189 B=224	α=0.10668378 β=0.00689633	R=405 P=37.979	R=08 P=4.6	A=0.10206 4 S=0.731718
G	44323.46	5.41	96.5	- = 47 + = 34	E=56630 A=1.278	E=56380 A=1.272	I=39.56 A=94.06	108,187, 256, 344	-0.168	E=184 B=213	α =0.00512976 β=0.000302109	R=386 P=19.2588	R=11 P=8.53	A=0.698514 S=0.6928
H	20073.00	6.08	99.8	- = 19 + = 17	E=17085 A=0.851	E=16960 A=0.845	I=27.20 A=80.28	75,179	-0.177	E=97 B=84	α=0.00030141 β =0.00027156	R=179 P=13.92	R=02 P=1.25	A=0.875572 S=0.626779
I	20054.96	6.08	99.9	- = 19 += 17	E=17085 A=0.852	E=1696 0 A=0.846	I=27.20 A=82.4	75,179	-0.166	E=94 B=87	α=0.00024775 β=0.00025249	R=168 P=12.1	R=13 P=8.4	A=0.87581 1 S=0.739278
J	20036.95	5.40	100.1	- = 21 + = 17	E=15595 A=0.778	E=15470 A=0.772	I=25.56 A=79.7	75,113	-0.187	E=98 B=83	α=0.00050846 β =0.00026878	R=165 P=11.4	R=16 P=11.8	A=0.89207 3 S=0.706462
K	20147.13	5.57	100.1	- = 21 + = 17	E=17085 A=0.848	E=16960 A=0.842	I=26.10 A=82.4	75,179	-0.169	E=95 B=86	α=0.00091331 β =0.00040871	R=166 P=11.270	R=15 P=11.070	A=0.89638 7 S=0.57387
L	20211.22	5.42	100.0	- = 21 + = 17	E=18575 A=0.919	E=18450 A=0.913	I=30.38 A=80.2	75,179	-0,185	E=98 B=83	α=0.00073988 β=0.00045345	R=166 P=10.3	R=15 P=10.5	A=0.86365 7 S=0.563569
M	15953.22	6.07	100.2	- = 14 + = 12	E=16960 A=1.063	E=16960 A=1.063	I=23.27 A=74.00	.	-0.135	E=79 B=66	α=0.00516498 β=0.00154569	R=132 P=10.88	R=13 P=9.42	A=0.926676 S=0.606109
N	42352.68	5.09	97.2	- = 41 + = 27	E=58370 A=1.378	E=57870 A =1.366	I=35.27 A=94.17	15,33,118, 148, 284,295	-0.030	E=174 B=205	.	R=370 P=21.0489	R=09 P=8.02	A=0.575821 S = 0.997670
O	28741.70	6.17	99.0	- = 27 + = 24	E=22710 A=0.790	E=2246 0 A=0.781	I=38.52 A=88.1	176,178, 208, 219	-0.076	E=118 B=140	α=0.001160075 β=0.000830566	R=254 P=15.468	R=04 P=2.3	A=0.83360 1 S=0.662839
P	20117.15	5.8	100.2	- = 20 + = 17	E=17085 A=0.849	E=1696 0 A=0.843	I=25.22 A=83.48	72,179	-0.139	E=96 B=8 5	α=0.00081903 β=0.00035442	R=166 P=10.	R=15 P=11. 0	A=0.90273 S=0.621267
Q	30767.09	7.7	98.5	- = 27 + = 28	E=37595 A=1.222	E = 37470 A= 1.218	I=36.13 A=81	175,278	-0.210	E = 170 B = 110	α=0.0032326 β=0.0058584	R=260 P=30	R=20 P=12. 5	A=0.92169 8 S=0.881002
R	37484.99	4.95	96.6	- = 41 += 25	E=43805 A=1.169	E=4343 0 A=1.159	I=41.8 A=91.9	76,103, 305, 328	-0.111	E=140 B=193	α=0.00823176 β=0.00249368	R=331 P=19.284	R=02 P=1.1 8	A=0.64600 7 S=0.990986
S	51649.00	9.26	99.9	- = 53 += 67	E=50100 A=0.970	E=4985 0 A=0.965	I=31.90 A=92.2	189,206, 421, 435	-0.277	E=207 B=242	α=0.0847252 β=0.0106443	R=441 P=27.5	R=08 P=5.7	A=0.63760 6 S=0.712093
T	35004.85	8.94	99.6	- = 39 + = 45	E=18700 A=0.534	E=1845 0 A=0.527	I=29.24 A=98.7	188,214	-0.275	E=148 B=157	α =0.0451376 β =0.00804493	R=288 P=18.378	R=17 P=14.5	A=0.45140 9 S=0.602309

(A)-Homosapiens;(B)-Rattus Norvegicus;(C)-Musmusculus;(D)-Nomascus-gabriellae;(E)-Rhinopithecus roxellana;(F)-Fukomys damarensis; (G)-Canislupus-familiaris;(H)-Panpiniscus;(I)-Pantroglodytes;(J)-Pongopygmaeus;(k)-Nomascus-leucogenys;(L)-Gorilla-Gorilla;(M)Hylobates pileatus;(N)-Phoca vitulina;(O)-Sapajus apella ; (P)-Hylobates moloch; (Q)-Piliocolobus tephrosceles;(R)-Suricatta-Suricatta; (S)-Microtus Ochrogaster;(T)-Urocitellus parryii.

PART-2:

Writing a code for developing database driven application.The main reason to develop this code is to secure the above tabulated data using user name and password.In the above table we have collected data for only limited organisms.If this work is continued in the future then to safe guard the data we have to develop application through NET-beans IDE using MySQL database.But in this work we restricted ourselves in writing a code. We ended with a error "Reached end of the file while parsing".According to the Net- informations,founded by raps mk this type of the error can be easily avoided by using a code editor like **NETBEANS** or **ECLIPSE** . Using these IDE's, you can autoformat your code by pressing Alt+Shift+F . This will indent your code properly and align matching braces with the control structure (loop, if, method, class) that they belong to.

https://www.docdroid.net/Mg3c0ay/writing-a-error-free-code-for-establishing-database-driven-application-with-mysql-database-pdf

PART-3:

We are aware of many preditors online for studying the protein properties in advanced way.In this we are presenting P^3-Protein Property Predictor which provides raw data i.e; numerical data,script/post script data.The following link gives the entire information about the usage of this predictor

https://www.docdroid.net/XqVX6h8/p3protein-property-predictor-usage-major-project-part-3-file-uploaded-by-divya-peddapalyam-pdf

IV. CONCLUSION:

On the whole this study of studying the pseudo protein(IRGM) properties suitable for numerous sophisticated bio medical applications in regenerative medicine,pharmacy etc.This properties also help us to represent protein samples for improving protein sub cellular localization prediction and membrane protein type prediction.This also tells us which organism pseudo protein induce very low inflammation,cell and tissue biocomapatible,non toxic,Enzymatic biodegradable,versatile with a wide range of the material and biological properties at a basic level..Biomedical applications range from wound care,healing,Drug delivary,scaffolds for tissue engineering,diagnostics and coating for medical devices.

V. REFERENCES:

1. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, et al. (2012) Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 142: 46–54**. [PubMed] [Google Scholar]**

2. Packey CD, Sartor RB (2008) Interplay of commensal and pathogenic bacteria, genetic mutations, and immunoregulatory defects in the pathogenesis of inflammatory bowel diseases. J Intern Med 263: 597–606. **[PubMed] [Google Scholar]**

3. Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, et al. (2001) A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. Nature 411: 603–6**. [PubMed] [Google Scholar]**

4. Hampe J, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, et al. (2001) Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. Lancet 357: 1925–8**. [PubMed] [Google Scholar]**

5. Pugazhendhi S, Amte A, Balamurugan R, Venkataraman S, Ramakrishna BS (2008) Common NOD2 mutations are absent in Crohn's disease patients in India. Indian J Gastroenterol 27: 201–3. **[PubMed] [Google Scholar]**

6. Pugazhendhi S, Santhanam S, Venkataraman J, Creveaux I, Ramakrishna BS (2013) NOD2 gene mutations associate weakly with ulcerative colitis but not with Crohn's disease in Indian patients with inflammatory bowel disease. Gene 512: 309–13. **[PubMed] [Google Scholar]**

7. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, et al. (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012 491: 119–24. **[PMC free article] [PubMed] [Google Scholar]**

8. Billmann-Born S, Lipinski S, Böck J, Till A, Rosenstiel P, et al. (2011) The complex interplay of NOD-like receptors and the autophagy machinery in the pathophysiology of Crohn disease. Eur J Cell Biol 90: 593–602. **[PubMed] [GoogleScholar]**

9. Parkes M, Barrett JC, Prescott NJ, Tremelling M, Anderson CA, et al. (2007) Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet 39: 830–832. **[PMC free article] [PubMed] [Google Scholar]**

10. Brest P, Lapaquette P, Souidi M, Lebrigand K, Cesaro A, et al. (2011) A synonymous variant in IRGM alters a binding site for miR- 196 and causes deregulation of IRGM-dependent xenophagy in Crohn's disease. Nat Genet 43: 242–5**. [PubMed] [Google Scholar]**

11. Taylor GA (2007) IRG proteins: key mediators of interferon-regulated host resistance to intracellular pathogens. Cell Microbiol 9: 1099–1107. **[PubMed] [Google Scholar]**

12. Singh SB, Davis AS, Taylor GA, Deretic V (2006) Human IRGM induces autophagy to eliminate Intracellular mycobacteria. Science 313: 1438–1441. **[PubMed] [Google Scholar]**