ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Homology Modeling Studies of Ras Protein in Cancer.

SK Salamuddin[1], Ch Venkata Rami Reddy[2], and A Ranganadha Reddy*

[1]School of Information Technology Vignan University, Guntur, Andhra Pradesh 522 213, India.
[2]Department of Computer Science Engg. Vignan University, Guntur, Andhra Pradesh 522 213, India.
*School of Biotechnology Vignan University, Guntur, Andhra Pradesh 522 213, India.

ABSTRACT

Cancer is a class of disease in which a group of cells display uncontrolled growth, invasion, and sometimes metastasis. The Ras GTPase proteins and their down-stream effectors regulate specific intracellular signalling pathways involved in numerous biological processes. Their actions directly influence progression through the cell cycle and the delicate balance of pro- and anti-apoptotic factors. The variety of functions controlled by Ras, and the emerging evidence indicating that aberrations in Ras as well as at multiple points in downstream signalling pathways contribute to tumourigenesis, suggest that Ras signal transduction mechanisms have significant potential as anti-cancer therapeutic targets. The RAS protein controls signaling pathway are major player in cell growth, its regulation and malignant transformation. Any activation in RAS brings alteration in upstream or downstream signaling component.Ras-specific guanine nucleotide-releasing factor 2 is a protein that in humans is encoded by the RASGRF2 gene. RAS (MIM 190020) GTPases cycle between an inactive GDP-bound state and an active GTP-bound state. Guanine-nucleotide exchange factors (GEFs), such as RASGRFs, stimulate the conversion of the GDP-bound form into the active form. Considering the importance and lack of specific structure of Ras protein we have modeled the three dimensitional (3D) structure by homology modeling method using Mol soft and model verification is carried out by using ramachandran plot analysis.
Keywords: Cancer, metastasis, Ras protein, homology modeling, Molsoft, Ramachandran plot

Corresponding author

ISSN: 0975-8585

INTRODUCTION

Cancer is a class of disease in which a group of cells display uncontrolled growth (division beyond the normal limits), invasion (intrusion on and destruction of adjacent tissues), and sometimes metastasis (spread to other locations in the body via lymph or blood). These three malignant properties of cancers differentiate them from benign tumors, which are self-limited, and do not invade or metastasize [1]. Most cancers form a tumor but some, like leukemia, do not. The branch of medicine concerned with the study, diagnosis, treatment, and prevention of cancer is oncology [2].

Cancers are caused by abnormalities in the genetic material of the transformed cells. These abnormalities may be due to the effects of carcinogens, such as tobacco smoke, radiation, chemicals, or infectious agents. Other cancer-promoting genetic abnormalities may randomly occur through errors in DNA replication, or are inherited, and thus present in all cells from birth. The heritability of cancers is usually affected by complex interactions between carcinogens and the host's genome. Genetic abnormalities found in cancer typically affect two general classes of genes. Cancer-promoting oncogenes are typically activated in cancer cells, giving those cells new properties, such as hyperactive growth and division, protection against programmed cell death, loss of respect for normal tissue boundaries, and the ability to become established in diverse tissue environments. Tumor suppressor genes are then inactivated in cancer cells, resulting in the loss of normal functions in those cells, such as accurate DNA replication, control over the cell cycle, orientation and adhesion within tissues, and interaction with protective cells of the immune system [3].

RAS PROTEIN

Ras genes are the most common targets for somatic gain-of-function mutations inhuman cancer. Recently, germline mutations that affect components of the Ras–Raf–mitogen-activated and extracellular-signal regulated kinase kinase (MEK)–extracellular signal-regulated kinase (ERK) pathway were shown to cause several developmental disorders, including Noonan, Costello and cardio-facio-cutaneous syndromes [4]. Ras specific guanine nucleotide-releasing factor 2 is a protein in humans that is encoded by the RASGRF2. RAS GTPases cycle between an inactive GDP-bound state and an active GTP-bound state. Guanine-nucleotide exchange factors (GEFs), such as RASGRFs, stimulate the conversion of the GDP-bound form into the active form.

p35/cyclin-dependent kinase 5 phosphorylation of ras guanine nucleotide releasing factor 2 (RasGRF2) mediates Rac-dependent Extracellular Signal-regulated kinase 1/2 activity, altering RasGRF2 and microtubule-associated protein 1b distribution in neurons. Cyclin-dependent kinase 5 (Cdk5) is a proline-directed kinase the activity of which is dependent on association with its neuron-specific activators, p35 and p39. Cdk5 activity is critical for the proper formation of cortical structures and lamination during development.

In the adult nervous system, Cdk5 function is implicated in cellular adhesion, dopamine signaling, neurotransmitter release, and synaptic activity. In addition, Cdk5 is also involved in "cross-talk" with other signal transduction pathways. To further examine its involvement in cross-talk with other pathways, we identified proteins that interacted with p35 using the yeast two-hybrid system. We report here that p35 associates with Ras guanine nucleotide releasing factor 2 (RasGRF2) in coimmunoprecipitation and colocalization studies using transfected cell lines as well as primary cortical neurons.Calmodulin-Independent Coordination of Ras and Extracellular Signal-Regulated Kinase Activation by Ras-GRF2 as shown in Figure-1.

Ras-GRF2 (GRF2) is a widely expressed, calcium-activated regulator of the small-type GTPases Ras and Rac. It is a multidomain protein composed of several recognizable sequence motifs in the following order (NH2 to COOH): pleckstrin homology (PH), coiled-coil, ilimaquinone (IQ), Dbl homology (DH), PH, REM (Ras exchanger motif), PEST/destruction box, Cdc25. The DH and Cdc25 domains possess guanine nucleotide exchange factor (GEF) activity and interact with Rac and Ras, respectively. The REM-Cdc25 region was found to be sufficient for maximal activation of Ras in vitro and in vivo caused Ras and extracellular signal-regulated kinase (ERK) activation independent of calcium signals, suggesting that, at least when expressed ectopically, it contains all of the determinants required to access and activate Ras signaling [5].

ISSN: 0975-8585

COMPUTATIONAL TOOLS

All calculations were carried out in Maestro v9.2 installed in Cadd-WS3 machine under 64-bit centos operating system placed in CADD department, Institute of Life Sciences. The machine was built up with:

A) 4 cores and 8 processers with Intel Xenon CPU E5620 @ 2.40GHZ
B) 16 GB RAM
C) NVidia Qudvo FX3800 Graphical Process Unit (GPU)
D) The PROCHECK analysis provides an idea of the stereo chemical quality of all protein chains in a given PDB structure. They highlight regions of the proteins which appear to have unusual geometry and provide an overall assessment of the structure as a whole.

E) Other Servers

- Primary sequence of the Ras Protein was retrieved from Swiss Prot (accession number **O14827**) from the ExPASy (Expert Protein Analysis System) proteomics serves of the Swiss Institute of Bioinformatics.
- Homology search for Ras protein was carried out using BLAST software.

F) The crystal structure for Ras protein (PDB ID: **2IJE**) was obtained from PDB database RCSB.

DATABASES

Database: A database is a collection of information that is organized so that it can easily be accessed, managed, and updated.

Data mining: Data mining [6] or knowledge discovery is the computer-oriented process of digging and analyzing large volumes of data and finally extracting the meaning of the data. Applications of data mining to bioinformatics include gene finding, protein function domain detection, function motif detection, protein function inference, disease diagnosis, disease prognosis, disease treatment optimization, protein and gene interaction network reconstruction, data cleansing, and protein sub-cellular location prediction.

Swiss-Prot: Swiss-Prot was was created in 1986 by Amos Bairoch and developed by the Swiss Institute of Bioinformatics and subsequently developed by Rolf Apweiler at the European Bioinformatics Institute. [7-9] Swiss-Prot aimed to provide reliable protein sequences associated with a high level of annotation a minimal level of redundancy and high level of integration with other databases.

Protein Data Bank: PDB consists of 3D (Three Dimensional) data of experimentally determined structures of proteins and nucleic acids [10] established at Brookhaven National Laboratory [11]. The archive is managed by the Worldwide Protein Data Bank organization (wwPDB), whose mission is to ensure that a single, global PDB data archive is and will remain freely and publicly available [12].

SOFTWARES USED

BIOEDIT: BioEdit is a user friendly biological sequence alignment editor and sequence analysis program, for windows 95/95/NT systems [13]. In-color alignment and editing with separate nucleic acid and amino acid color tables and full control over background colors. BioEdit can be downloaded from the following Servers for sequence analysis. .http://www.mbio.ncsu.edu/BioEdit/bioedit.html

MOLSOFT: ICM MOLSOFT [14] algorithm was adopted for comparative modeling which provides an accurate and efficient module to build loops and side chains found non-identical in sequence. ICM molsoft algorithm contains robust modeling tools and high levels of accuracy with fast model building.

ISSN: 0975-8585

TOOLS

Basic Local Alignment Search Tool (BLAST): Nowadays Similarity searching, including sequence comparison, is one of the principal techniques used by computational biologists and has found widespread use among biologists in general. The most popular tool for this purpose is BLAST [15] (**Basic Local Alignment Search Tool)** which performs comparisons between pairs of sequences, searching for regions of local similarity. NCBI BLAST is available from the NCBI [16] **(National Center for Biotechnology Information**).

Self-Optimized Prediction Method with Alignment (SOPMA): It is a self optimizing prediction method of alignment and is used for prediction of secondary structure of proteins. This method calculates the content of alpha helix, beta sheets, turns, random coils and extended strands. SOPMA [17] is available online on http://www.expasy.org [18].

WEB BASED SERVERS

Real Space Automated Conformer Generation (RAPPER):RAPPER [19] is an ab initio conformational search algorithm for restraint-based protein modeling. It has been used for all-atom loop modeling, whole protein modeling under limited restraints comparative modeling, ab initio structure prediction, structure validation and experimental structure determination with X-ray and nuclear magnetic resonance spectroscopy. Web interfaces are available on this website for Ramachandran plot analysis.

METHODOLOGY

To build a homology model of Ras protein using ICM MOLSOFT software.

Protein Preparation (Prep Wiz)

A typical PDB structure consists of heavy atoms, waters, cofactors, metal ions and can be multimeric. The structure generally has no information on bond orders, topologies, or formal atomic charges. So, 2IJE (from the PDB) must be prepared by using protein preparation wizard (PrepWiz) of Schrödinger software. Protein preparation ensures that the 2IJE protein structure was properly assigned with bond orders and correct number of hydrogens to make the structure compatible with the OPLS (Optimized Potential for Liquid Simulations) forcefields [20] and the process of building the homology model MOLSOFT is used.

Homology modeling

The steps to creating a homology model are as follows:

- Identify homologous proteins and determine the extent of their sequence similarity with one another and the unknown.
- Align the sequences.
- Identify structurally conserved and structurally variable regions s
- Generate coordinates for core (structurally conserved) residues of the unknown structure from those of the known structure(s).
- Generate conformations for the loops (structurally variable) in the unknown structure.
- Build the side-chain conformations.
- Refine and evaluate the unknown structure.

Validation of Modeled Protein using RAPPER

As the loop of the structure is built up and the terminals are removed, the modeled structure is now given for validation through RAPPER. The structure is given in .pdb form to it to analyze the structure and to generate the Ramachandran plot of the structure to find the validity of the structure. This is again also carried out after minimization of the structure is done.

ISSN: 0975-8585

RESULTS AND DISCUSSION

Sequence Retrieval and Data Collection

The details of the Ras protein like amino acid lengths of N-and C-terminus, helix lengths etc were obtained from Protein knowledgebase UniProtKB/Swiss-Prot database with accession number **O14827.** The details of sequence annotation are shown below.

Query sequence of Ras Protein

```
>sp|O14827|RGRF2_HUMAN Ras-specific guanine nucleotide-releasing factor 2 OS=Homo
sapiensGN=RASGRF2PE=1
SV=2SAMELAEQITLLDHVIFRSIPYEEFLGQGWMKLDKNERTPYIMKTSQHFNDMSNLVASQIMNYADVSSRANAIEKWVAVADICRCLHN
YNGVLEITSALNRSAIYRLKKTWAKVSKQTKALMDKLQKTVSSEGRFKNLRETLKNCNPPAVPYLGMYLTDLAFIEEGTPNFTEEGLVNFSK
MRMISHIIREIRQFQQTSYRIDHQPKVAQYLLDKDLIIDEDTLYELSLKIEPR
```

relative positions of amino acids, hydrophilic residues located in core region. Parker-HPLC program shown in **Figure-6** is used to identify hydrophilic regions, the plot reliably predict 2 hydrophilic regions above the sliding window, which are located in the three dimensional structure at amino acid positions 28-150 and 160-210 residues are hydrophilic regions in Ras protein.

Secondary structure prediction by SOPMA

The secondary structures of proteins are the regularly repeating local structures stabilized by hydrogen bonds. The most common examples are the alpha-helix and beta-sheet.Secondary structure analysis of Ras protein was carried out using SOPMA as shown in **Figure-7**.SOPMA view predicted the following secondary structures.**Table-1** shows percentage of Alpha helix (Hh) is 68.24%%, percentage of 3_{10} helix (Gg) is 0.00% , percentage of Pi helix (Ii) is 0.00% , percentage of Beta bridge (Bb) is 0.00% , percentage of Extended strand (Ee) is 03.43% , percentage of Beta turn (Tt) is 1.72% , percentage of Bend region (Ss) is 0.00% , percentage of Random coil (Cc) is 26.61% percentage of Ambigous states is 0.00% and Other states is 0.00%.

Secondary Structure prediction by SOPMA

SOPMA result for : UNK_327350

Abstract Geourjon, C. & Deléage, G., SOPMA: Significant improvement in protein secondary structure prediction by consensus prediction from multiple alignments.,
681-684

View SOPMA in: [AnTheProt (PC) , Download...] [HELP]

```
        10         20         30         40         50         60         70
         |          |          |          |          |          |          |
SAMELAEQITLLDHVIFRSIPYEEFLGQGWMKLDKNERTPYIMKTSQHFNDMSNLVASQIMNYADVSSRA
chhhhhhhhhhhhhhhhhcccchhhhhhhhcccccccccchhhhhhhhhhhhhhhhhhhhhhcccchhh
NAIEKWVAVADICRCLHNYNGVLEITSALNRSAIYRLKKTWAKVSKQTKALMDKLQKTVSSEGRFKNLRE
hhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccchhhhhhhhhhccchhhhhhhhhhhhhhhhttchhhhhh
TLKNCNPPAVPYLGMYLTDLAFIEEGTPNFTEEGLVNFSKMRMISHIIREIRQFQQTSYRIDHQPKVAQY
hhhhcccccceeeeehhhheeecccccccccchhhhhhhhhhhhhhhhhhhtcccccccccchhhhh
LLDKDLIIDEDTLYELSLKIEPR
hhhhccccchhhhhhhhhhcct
```

Sequence length : 233

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between protein or nucleotide sequences. The program compares nucleotide or protein sequences to sequence in a database and calculates the statistical significance of the matches. The target sequence i.e., Ras (UniProt ID: **O14827**) was searched against the protein database by using BLAST tool. From the BLAST results as shown in **Figure-8** we observed that four proteins (PDB IDs: 2IJE-A, 1XDV-A, 2IIO-A, 1BKD-A) are showing the maximum identity with the target sequence. Among the four proteins obtained, **2IJE-A** selected for further proceedings.

ISSN: 0975-8585

Three-Dimensional Structure Prediction of Ras Protein by MOLSOFT Software

Three dimensitional structure of Ras Protein (**O14827**) as shown in **Figure-9** was predicted by using the tool MOLSOFT ICM by taking 2IJE-A as template which was obtained through BLAST results by taking **O14827** as query sequence and performing Protein BLAST against the protein sequence database.

Assessment of the Homology Model of Ras Protein

The validation of the final model was carried out using Ramachandran plot computed with PROCHECK, program by checking the detailed residue-by-residue stereo-chemical quality of a protein structure. The PROCHECK is used for stereo chemical assessment of the model. The criteria for analysis of stereochemistry of the model includes,

1) Main chain conformation in acceptable regions of the Ramachandran plot.
2) Planar peptide bonds.
3) Side chain conformations that correspond to those in rotamer library.
4) Hydrogen bonding of polar atoms if they are buried.
5) No bad atom-atom contacts.
6) No holes inside the structure.

Ramachandran Plot analysis by RAPPER

A Ramachandran plot (also known as a Ramachandran map or a Ramachandran diagram or a [Φ,Ψ] plot), developed by Gopalasamudram Narayana Ramachandran and Viswanathan Sasisekharan is a way to visualize dihedral angles Ψ and Φ of amino acid residues in protein structure. It shows the possible conformations of Φ and Ψ angles for a polypeptide. Hence, Ramachandran plot is a useful way of assessing the stereo chemical quality of a protein structure. The final refined model was further assessed by RAPPER server (ramachandran plot analysis), and the results show that Ras protein had 229 (97.4%) residues in favoured region against (~98.0% expected), 06 (2.6%) residues in allowed region against (~2.0% expected) and 0 (0.0% residues in outlier region as shown in Figure-10 which shows that the final model is reliable.

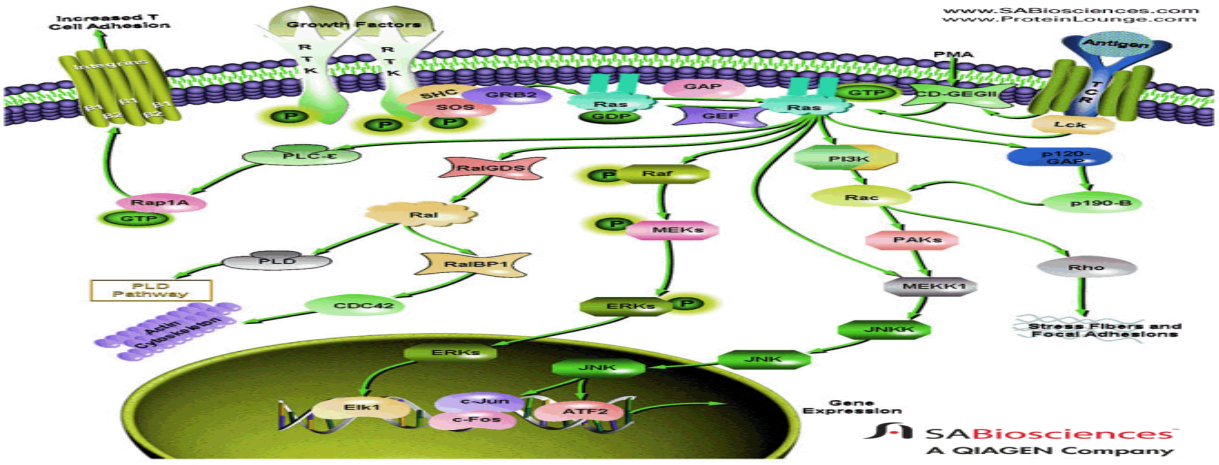


Figure-1: Signal Transduction Pathway Involving Ras Protein

ISSN: 0975-8585

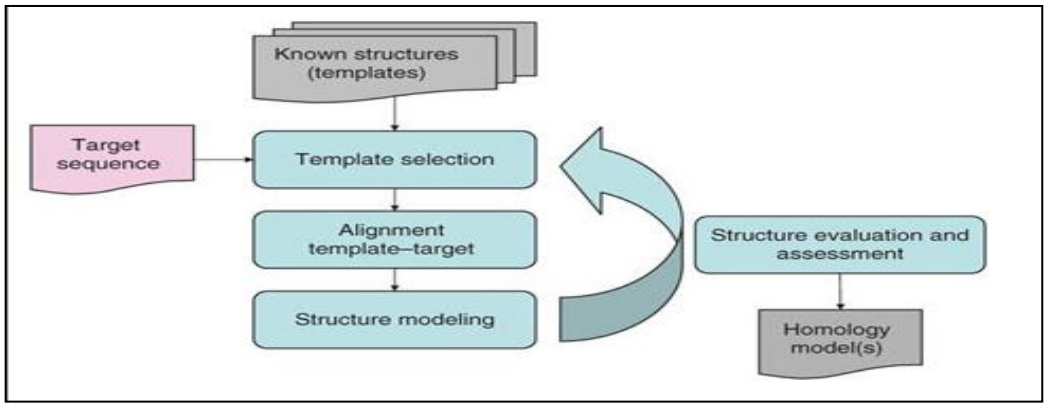


Figure -2: The four main steps of comparative protein structure modeling: template selection, target–template alignment, model building and model quality evaluation [14]

Protein: sp|O14827|RGRF2_HUMAN Ras-specific guanine nucleotide-releasing factor 2 OS=Homo sapiens GN=RASGRF2 PE=1 SV=2
Length = 233 amino acids
Molecular Weight = 27031.73 Daltons

Amino Acid		Number	Mol%
Ala	A	15	6.44
Cys	C	3	1.29
Asp	D	12	5.15
Glu	E	17	7.30
Phe	F	8	3.43
Gly	G	7	3.00
His	H	5	2.15
Ile	I	18	7.73
Lys	K	17	7.30
Leu	L	24	10.30
Met	M	9	3.86
Asn	N	13	5.58
Pro	P	8	3.43
Gln	Q	11	4.72
Arg	R	13	5.58
Ser	S	16	6.87
Thr	T	13	5.58
Val	V	11	4.72
Trp	W	3	1.29
Tyr	Y	10	4.29

Figure-3: The above graph is showing amino acid composition of Ras-specific guanine nucleotide releasing factor 2 amino acid LEU is present highest percentage compared to other residues and the molecular weight of the protein is 10.30

Prediction of Hydrophobicity

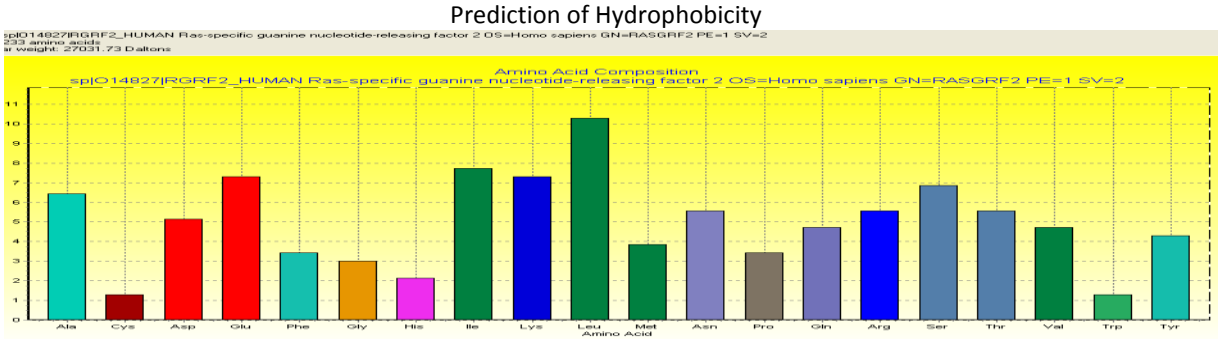


Figure-4: The above graph is showing amino acid composition of Query sequence,The amino acid LEU is present highest percentage compared to other residues.

ISSN: 0975-8585

Helical wheel diagram

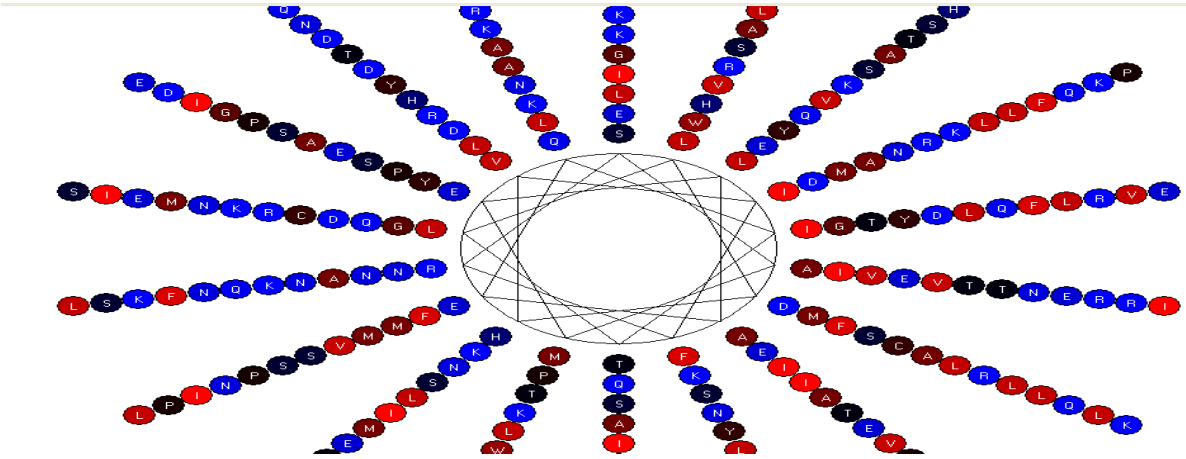


Figure-5: Wheel plot of the Ras-protein .the plot Shown obtained using Genetic Computer group HELICAL WHEEL PROGRAM, the diagram shows the relative positions of amino acids, hydrophobic residues located in core region.

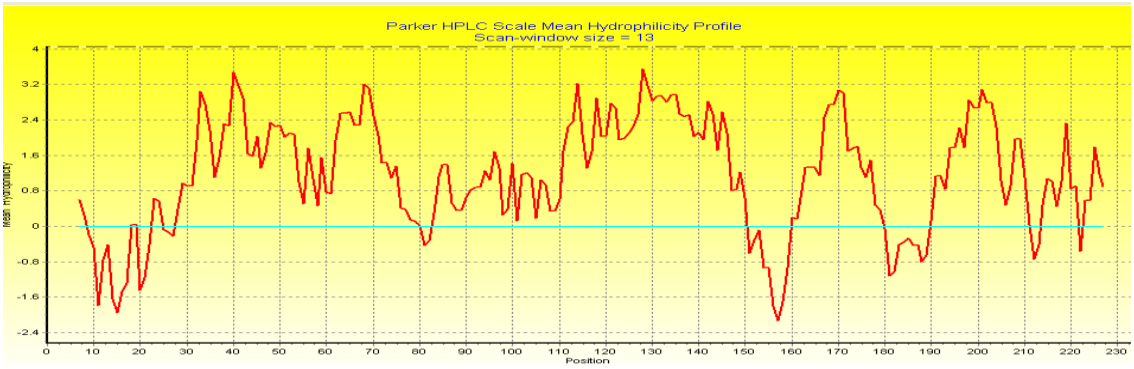
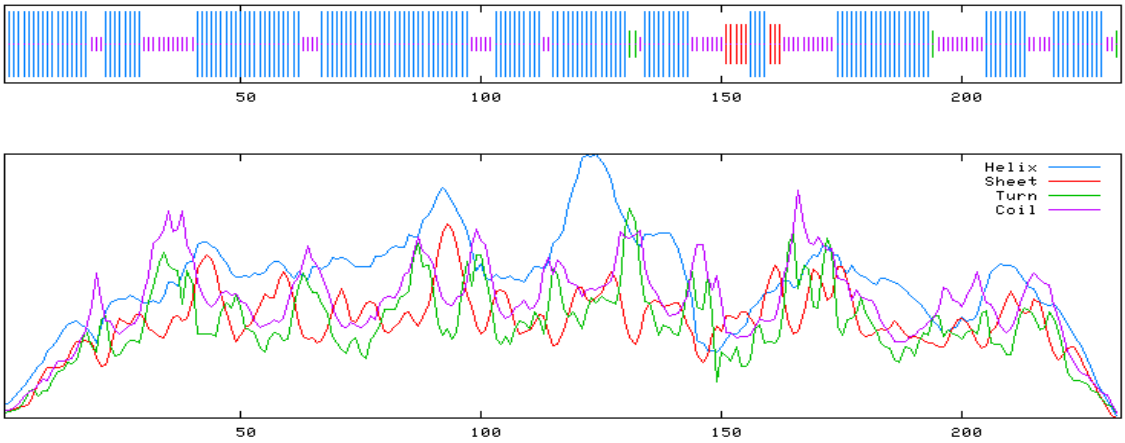


Figure-6: Hydrophilicity profile of sp_0014827_RGRF2_HUMAN above the slide window is indicating 28-150 and 160-210 residues are hydrophilic regions.

Table-1: Secondary structures predicted by SOPMA

Protein structure, Unit	No. of amino acids	Percentage of structural, Unit
Alpha helix((Hh)	159	68.24%
3_{10} helix (Gg)	0	0.00%
Pi helix (Ii)	0	0.00%
Beta bridge (Bb)	0	0.00%
Extended strand(Ee)	08	03.43%
Beta turn (Tt)	04	1.72%
Bend region (Ss)	0	0.00%
Random coil (Cc)	62	26.61%
Ambigous states	0	0.00%
Other states	0	0.00%

ISSN: 0975-8585




```
Parameters :
   Window width          :  17
   Similarity threshold  :   8
   Number of states      :   4
```

Figure-7: Secondary structure analysis of Ras Protein from SOPMA server

BLAST

Figure-8: Blast search result for sp_0014827_RGRF2_HUMAN as a query

ISSN: 0975-8585

Homology modeling by **MOLSOFT**

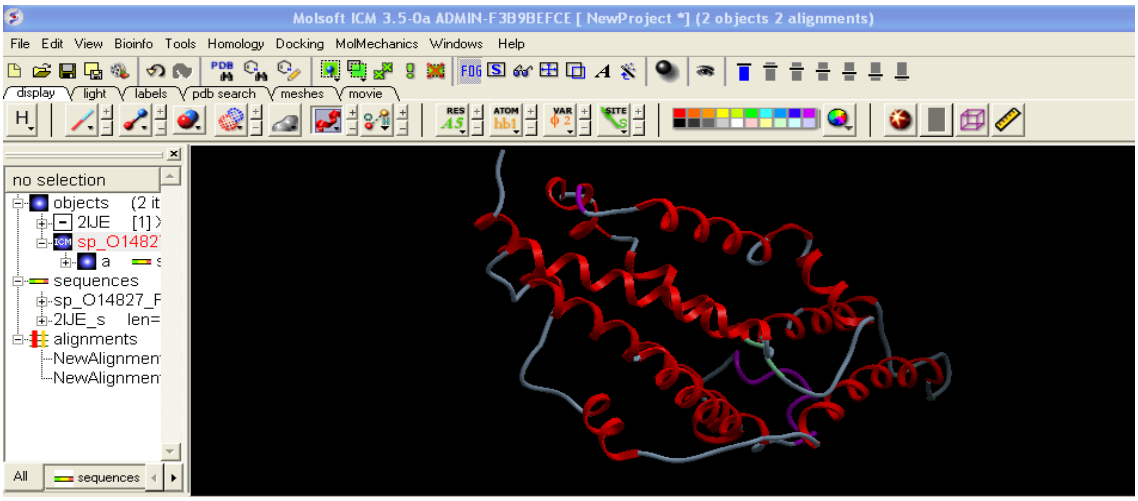
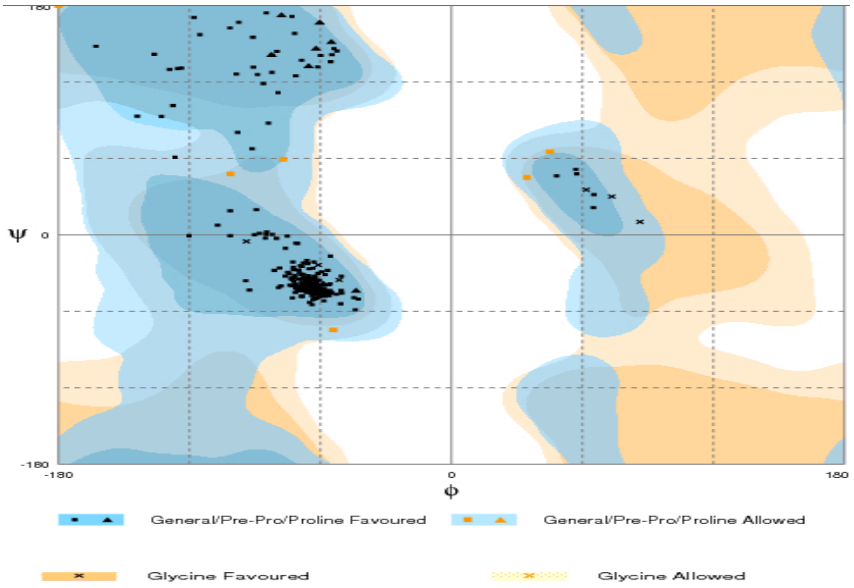
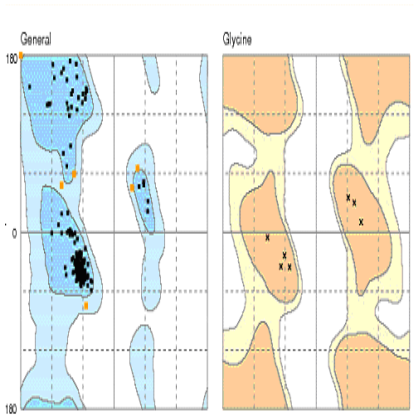
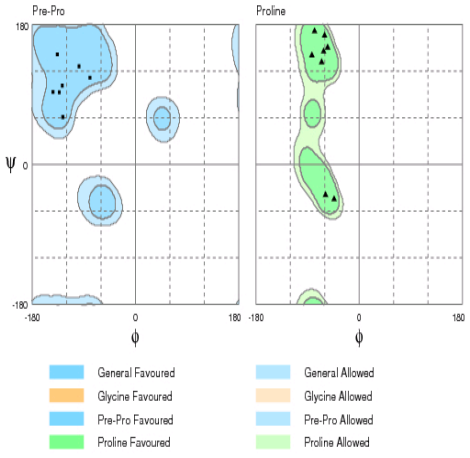


Figure-9: Shows Query (HUMAN Ras-specific guanine nucleotide-releasing factor 2) structure was modeled by using ICM mol soft by taking 2IJE as template.







ISSN: 0975-8585

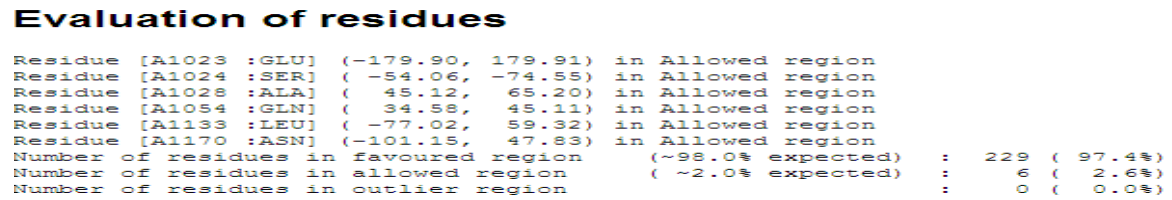


Figure-10: Protein validation study by RAPPER Server

CONCLUSION

Present study, we have constructed a 3D model of Ras Guanine nucleotide releasing factor 2 protein which is well known for its down-stream effectors, which regulate specific intracellular signaling pathways involved in numerous biological process using Mol soft software. Obtained a refined model after energy minimization. The final refined model was further assessed by RAPPER server (ramachandran plot analysis), and the results show that this model is reliable.

REFERENCES

[1] www.communityoncology.org
[2] Siegel R, et al. A Cancer Journal for Clinicians 2012;62:220–241.
[3] www.imtech.res.in
[4] Schubbert S, Shannon K, Bollag G Nat Rev Cancer 2007;7(4):295-308.
[5] De Hoog Carmen et al. Mol Cell Biol 2000:20(8):2727-2733.
[6] TC Venkateswarulu et al. Res J Pharm Biol Chem Sci 2014:5(3): 1417–1429.
[7] Bairoch A, Apweiler R. Nucleic Acids Res 1996:24 (1)21–25.
[8] Bairoch A. Bioinformatics 2000:16 (1): 48–64.
[9] Séverine Altairac. Naissance d'une banque de données: Interview du prof. Amos Bairoch". Protéines à la Une, August 2006. ISSN 1660-9824.
[10] Berman H. Acta Crystallogr A 2008;64:88–95.
[11] Bernstein FC, et al. J Mol Biol 1977;112:535–542.
[12] Berman HM, Henrick K, Nakamura H. Nat Struct Biol 2003;10:980.
[13] Hall TA . Nucl Acids Symp Ser 1999;41:95-98.
[14] A Ranganadha Reddy et al. Research Journal of Pharmacy and Technology, 2014;7(3):376-388.
[15] A Ranganadha Reddy, Sreedhara R Voleti, Ch Lakshmi Padma. Res J Pharm Biol Chem Sci 2013;4(4):151-168.
[16] Alexander Pertsemlidis and John W Fondon. Gen Biol 2001;2.
[17] Prashant V Thakare, et al. Global J Mol Sci 2010;5(1):30-36.
[18] http://www.expasy.org
[19] de Bakker PI, et al. Proteins 2003;51:21-40.
[20] Laurence J. Miller, Quan Chen, Polo C.-H. Lam, Delia I. Pinon, Patrick M. Sexton, Ruben Abagyan, and Maoqing Dong. J Biol Chem 2011;286:15895– 15907.