



IJIRR

International Journal of Information Research and Review
Vol. 04, Issue, 12, pp.4957-4962, December, 2017



RESEARCH ARTICLE

COMPARITIVE PROTEIN MODELING OF TUMOR SUPPRESSOR PROTEIN, RASSF1

*Udhayakumar, R. and Dr. Sudarsanam, D.

Department of Advanced Zoology and Biotechnology, Loyola College, Chennai-600034, Tamilnadu, India

ARTICLE INFO

Article History:

Received 22nd September, 2017
Received in revised form
19th October, 2017
Accepted 21st November, 2017
Published online 30th December, 2017

Keywords:

Lung cancer, Bioinformatics tools NCBI, MUSTER.

Copyright © 2018, Dr. Sudarsanam and Udhayakumar. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Lung cancer is a leading cause of cancer death worldwide. Although smoking remains the predominant cause of lung cancer, lung cancer in never smokers is an increasingly prominent public health issue.

INTRODUCTION

Lung cancer develops when normal lung cells sustain genetic damage that eventually leads to uncontrolled cell proliferation. Like all cancers, lung cancer cells have the ability to invade neighboring tissues and spread or metastasize to distant parts of the body. Left untreated, lung cancer eventually causes death. Lung cancer is sometimes referred to as bronchiogenic cancer or bronchiogenic carcinoma. The word bronchiogenic means originating from the bronchi, the airways of the lungs. Most lung cancers begin in the cells lining the bronchi of the lungs. There are two main types of lung cancer, and they are treated differently. Learning about your particular type of lung cancer can help you talk with your health care providers about your treatment choices. Lung cancer kills more people in the United States every year than breast, colon, and prostate cancer combined. The American Cancer Society estimates there will be 172,570 17 new cases of lung cancer in the United States in 2005, and 163,510 lung cancer deaths. This means that every day of the year in this country, approximately 470 people are diagnosed with lung cancer and 450 people die of the disease. Lung cancer is the leading cause of cancer deaths not only in the United States, but worldwide. The World Health Organization (WHO) reports that over 1.1 million people die of lung cancer each year. This number increases every year. As a result, WHO has identified lung cancer as one of the major problems facing the world in this new century (Silverstri and Jett, 2010). Lung Cancer Trends Lung cancer is a relatively new problem for human beings. In the early 1900's, lung cancer was

extremely rare. Although people have used tobacco for centuries, until the early 1900's, it was most often smoked with a pipe, in a cigar form, or chewed. The machines to massproduce cigarettes were invented in the 1880's, but it was not until after World War I that cigarette smoking became prevalent in the United States. During the war, American soldiers were given free cigarettes donated by tobacco companies to the U.S. military. Many soldiers left the military with a lifelong addiction to the nicotine in tobacco products. At that time, the medical community did not realize the long-term impact of cigarette smoking on health. There is usually a 20-30 year lag time between the onset of smoking and the development of lung cancer. In the 1930's, doctors started to note a dramatic rise in the incidence of lung cancer. By 1950, several studies reported the apparent link between rising lung cancer incidence and the new wave of cigarette smoking. (American Cancer Society, 2012; Ettinger, 2008; U.S. Department of Health and Human Services 2010; Aldington *et al.*, 2008; Temel *et al.*, 2010) In 1964, Dr. Luther L. Terry released the first Surgeon General's Report on Smoking and Health that definitively established the role of smoking as a contributing factor in the development of lung cancer and several other diseases. Until recently, lung cancer was viewed primarily as a man's disease. When lung cancer incidence first began to skyrocket, most cases were observed in men. This was because when cigarette smoking first became prevalent in this country, it was primarily among men. Over time, advertising campaigns aimed at women and other factors led to cigarette smoking eventually becoming almost as common in women as in men. The result of this unfortunate 18 trend has been an ever-increasing incidence of lung cancer among women. Death rates from lung cancer in women increased an astonishing

*Corresponding author: Udhayakumar, R.

Department of Advanced Zoology and Biotechnology, Loyola College, Chennai-600034, Tamilnadu, India.

550% between 1962 and 1992. In 1987, deaths from lung cancer among women surpassed breast cancer deaths. Today, for every three women that die of breast cancer, five women die of lung cancer. (National Cancer Institute, 2011) Clearly, lung cancer is no longer a man's disease. RISK FACTORS FOR DEVELOPING LUNG CANCER Lung cancer develops when the cells that line the lungs sustain genetic damage. Scientists have identified several different chemicals and environmental factors that are capable of causing the kind of genetic damage that can lead to lung cancer. Substances capable of producing cancerous changes in cells are called carcinogens. The majority of lung cancers occur in people who are either current or former smokers. While the relationship between smoking and lung cancer is well-established, other factors also come into play. We know this because only about one out of every ten smokers develops lung cancer. Further, approximately one out of every six people who develops lung cancer never smoked.

MATERIALS AND METHODS

Databases used

NCBI
 PROCHEK
 3DPSSM
 MUSTER
 PUBMED
 STRING DATABASE

Tools used

SWISSPDB VIEWER
 MOLGERO MOLECULAR VIEWER
 DATABASES

NCBI

The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health. The NCBI is located in Bethesda, Maryland and was founded in 1988 through legislation sponsored by Senator Claude Pepper. The NCBI houses genome sequencing data in GenBank and an index of biomedical research articles in PubMed Central and PubMed, as well as other information relevant to biotechnology. All these databases are available online through the Entrez search engine. The NCBI is directed by David Lipman, one of the original authors of the BLAST sequence alignment program and a widely respected figure in Bioinformatics. He also leads an intramural research program, including groups led by Stephen Altschul (another BLAST co-author), David Landsman, and Eugene Koonin (a prolific author on comparative genomics).

PROCHECK

The PROCHECK analyses provide an idea of the stereochemical 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. The analyses can be generated by clicking on the PROCHECK button at the top of the PROCHECK summary page; you can get to this page from the

PDBsum page of the structure in question either by clicking on PROCHECK in the index on the left hand side of the page, or else by clicking on the Ramachandran plot on the right.

3DPSSM

The 3D-PSSM server is a designed to take a PROTEIN sequence of interest to you, and attempt to predict its 3-dimensional structure and its probable function. We have a library of known protein structures onto each of which your sequence is "threaded" and scored for compatibility. We use a variety of scoring components: 1D-PSSMs (sequence profiles built from relatively close homologues), 3D-PSSMs (more general profiles containing more remote homologues - see Methods), matching of secondary structure elements, and propensities of the residues in your query sequence to occupy varying levels of solvent accessibility

MUSTER

MUSTER (MUlti-Sources ThreadER) is a new protein threading algorithm to identify the template structures from the PDB library. It generates sequence-template alignments by combining sequence profile-profile alignment with multiple structural information. The outputs of the MUSTER server include:

- Top five template proteins and the query-template alignments
- Full-length models built by MODELLER

PUBMED

PubMed is a free database accessing the MEDLINE database of citations, abstracts and some full text articles on life sciences and biomedical topics. The United States National Library of Medicine (NLM) at the National Institutes of Health (NIH) maintains PubMed as part of the Entrez information retrieval system. Listing an article or journal in PubMed is not endorsement. In addition to MEDLINE, PubMed also offers access to

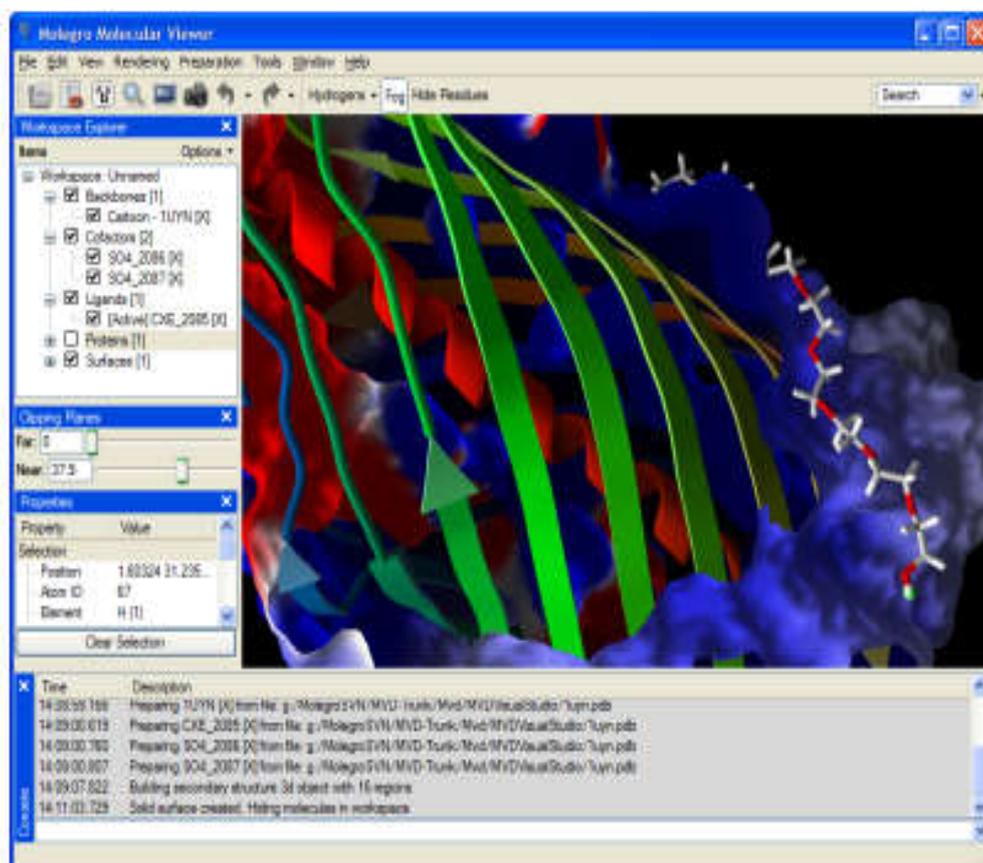
- OLDMEDLINE for pre-1966 citations. This has recently been enhanced, and records for 1951+, even those parts in the printed indexes, are now included within the main portion.
- Citations to all articles (even those that are out-of-scope, e.g., covering plate tectonics or astrophysics) from certain MEDLINE journals, primarily the most important general science and chemistry journals, from which the life sciences articles are indexed for MEDLINE.
- In-process citations which provide a record for an article before it is indexed with MeSH and added to MEDLINE or converted to out-of-scope status.
- Citations that precede the date that a journal was selected for MEDLINE indexing (when supplied electronically by the publisher).
- Some life science journals that submit full text to the PubMed Central digital library and may not have been recommended for inclusion in MEDLINE although they have undergone a review by NLM, and some physics journals that were part of a prototype PubMed in the early to mid-1990s.

Many PubMed citations contain links to full text articles which are freely available, often in PubMed Central. In late 2007, President George W. Bush signed the Consolidated Appropriations Act of 2007 (H.R. 2764) into law; this law included a provision requiring the NIH to modify its policies and require inclusion into PubMed Central complete electronic copies of their peer-reviewed research and findings from its funded research. This is the first time the US government has required an agency to provide open access to research and is an evolution from the 2005 policy, in which the NIH asked researchers to voluntarily add their research to PubMed Central. [Silverstri and Jett, 2010] With an effective date of 7 April 2008, the Department of Health and Human Services gave notice: "The Director of the National Institutes of Health shall require that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication: Provided, That the NIH shall implement the public access policy in a manner consistent with copyright law". The National Library of Medicine also leases the MEDLINE information to a number of private vendors such as Ovid and SilverPlatter – as well as many other vendors. PubMed has been available free on the Internet since the mid-1990s. Information about the journals indexed in PubMed is found in its Journals Database, searchable by subject or journal title, Title Abbreviation, the NLM ID (NLM's unique journal identifier), the ISO abbreviation, and both the print and electronic International Standard Serial Numbers (pISSN and eISSN). The database includes all journals in all Entrez databases. As of 28 January 2010, PubMed has approximately 19.5 million citations going back to the year 1865.

Tools

Swisspdb viewer

DeepView is a friendly but powerful molecular graphics program. It is designed for full compatibility with computing tools available from the Expert Protein Analysis System, or ExpASY* Molecular Biology Server in Geneva, Switzerland. While DeepView is simple to use for viewing structures and creating vivid illustrations, it also shines as an analytical tool. DeepView allows you to build models from scratch, simply by giving an amino-acid sequence. DeepView can find hydrogen bonds within proteins and between proteins and ligands. It allows you to examine electron-density maps from crystallographic structure determination, to judge the quality of maps and models, and to identify many common types of problems in protein models. It allows you to view several or many models simultaneously and superimpose them to compare their structures and sequences. It computes electrostatic potentials and molecular surfaces, and carries out energy minimization. For proteins of known sequence but unknown structure, DeepView submits amino acid sequences to ExpASY to find homologous proteins, onto which you can subsequently align your sequence to build a preliminary three-dimensional model. Then DeepView submits your alignment to ExpASY, where the SWISS-MODEL server builds a final model, called a homology model, and returns it directly to DeepView. Swiss-PdbViewer was created by Nicolas Guex, and is under continuing development by Nicolas, along with Alexandre Diemand, Manuel C. Peitsch, and Torsten Schwede of GlaxoSmithKline R&D and the Swiss Institute of Bioinformatics. Thanks to Glaxo and to Nicolas and his colleagues for making this marvelous tool available to everyone. The program has been described in these articles:



The main user interface

- Protein Modelling for All, N. Guex, A. Diemand and M. C. Peitsch, Trends Biochem. Sci. 24 (9), 364-367 (1999).
- SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, N. Guex and M.C. Peitsch, Electrophoresis, 18 (15), 14-23 (1997).
- Swiss-PdbViewer: A new fast and easy to use PDB viewer for the Macintosh, N. Guex, Experientia 52, A26 (1996).
- Swiss-PdbViewer: A Fast and Easy-to-use PDB Viewer for Macintosh and PC, N. Guex and M.C. Peitsch,, Protein Data Bank Quarterly Newsletter 77, 7 (1996).

Molegro molecular viewer

Molegro Molecular Viewer is a free cross-platform application for visualization of molecules and Molegro Virtual Docker results. Molegro Molecular Viewer is able to visualize most common molecular file formats (PDB, Mol2, SDF) as well as docking results from Molegro Virtual Docker.

The latest version of MMV is 2.0.0 (released January 13th, 2010).

Features at a glance

- Free!
- Cross-platform: Windows, Linux, and Mac OS X are supported.
- Share and view results from Molegro Virtual Docker docking runs.
- Imports and exports PDB, SDF, Mol2, and MVDML files.
- Built-in raytracer for high-quality images.
- Automatic preparation of molecules.
- Molecular surface and backbone visualization.
- Labels, sequence viewer and biomolecule generator.
- Cropping of molecules and clipping planes.
- Structural protein alignment

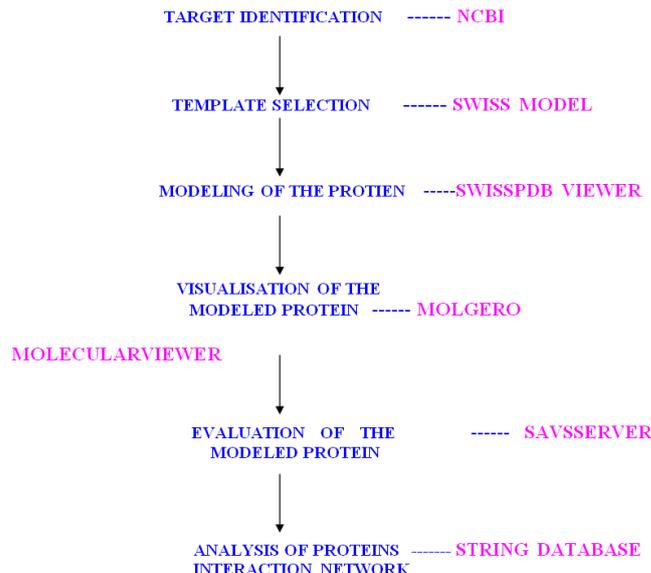
Identification of the template

Swiss-model template selection

ExPDB Sequences with high scorings

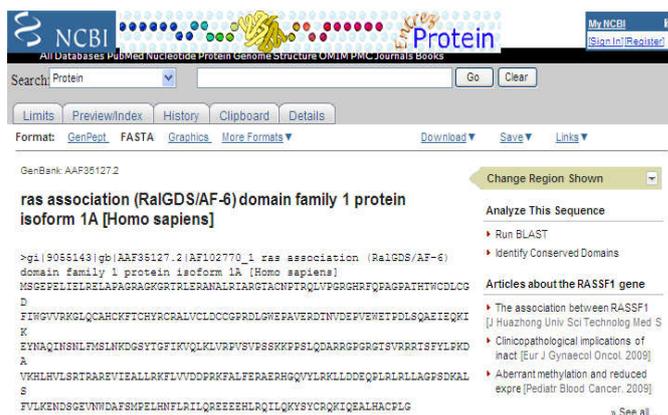
	download ExPDB	Blast Score	see	Exp.	Reso.	Parent PDB	Description
<input type="checkbox"/>	2mfX	2e-09	Detail	NMR,	99.9	2mf	C1 DOMAIN OF NORE1
<input type="checkbox"/>	1rfhA	2e-09	Detail	NMR,	99.9	1rfh	SOLUTION STRUCTURE OF THE C1 DOMAIN OF NORE1, A NOVEL RAS EFFECTOR
<input type="checkbox"/>	1ptrA	8e-07	Detail	X-RAY	2.20	1ptr	PROTEIN KINASE C DELTA CYS2 DOMAIN COMPLEXED WITH PHORBOL-13-ACETATE
<input type="checkbox"/>	1ptqA	8e-07	Detail	X-RAY	1.95	1ptq	PROTEIN KINASE C DELTA CYS2 DOMAIN
<input type="checkbox"/>	1y8fA	3e-06	Detail	NMR,	99.9	1y8f	SOLUTION STRUCTURE OF THE MUNC13-1 C1-DOMAIN
<input type="checkbox"/>	1tboA	1e-04	Detail	NMR,	99.9	1tbo	NMR STRUCTURE OF A PROTEIN KINASE C-G PHORBOL-BINDING DOMAIN, 30 STRUCTURES
<input type="checkbox"/>	1xa6A	1e-04	Detail	X-RAY	3.20	1xa6	CRYSTAL STRUCTURE OF THE HUMAN BETA2-CHIMAERIN
<input type="checkbox"/>	1tbnA	1e-04	Detail	NMR	99.9	1tbn	NMR STRUCTURE OF A PROTEIN KINASE C-G PHORBOL-BINDING DOMAIN, MINIMIZED AVERAGE STRUCTURE

METHODOLOGY



RESULTS

Fasata format of the target protein sequence using NCBI



```

• iptra EOP
PROTEIN KINASE C DELTA CYS2 DOMAIN COMPLEXED WITH PHORBOL-
13-ACETATE
MOL_ID: 1;
MOLECULE: PROTEIN KINASE C DELTA TYPE;
CHAIN: A;
EC: 2.7.1.-;
ENGINEERED: YES
MOL_ID: 1;
ORGANISM_SCIENTIFIC: MUS MUSCULUS;
ORGANISM_COMMON: MOUSE;
EXPRESSION_SYSTEM: ESCHERICHIA COLI
G.ZHANG,J.H.HURLEY
EQS assignment: HOMO DIMERIC

Length = 50

Score = 52.7 bits (124), Expect = 8e-07
Identities = 20/50 (40 Positives = 30/50 (60

Query: 52 HRFQFAGFATHWCDLQDFIWGVRKGLQCAHCKFTCHYRCRALVCLDC 101
          HRF          T CD CG WG V GL C C H CR V C
Sbjct: 1 HRFKVVYVMSPTFCDCGSLWLVKQGLKCEDCGMNVHKKCEKVKANLC 50
    
```

Plot statistics

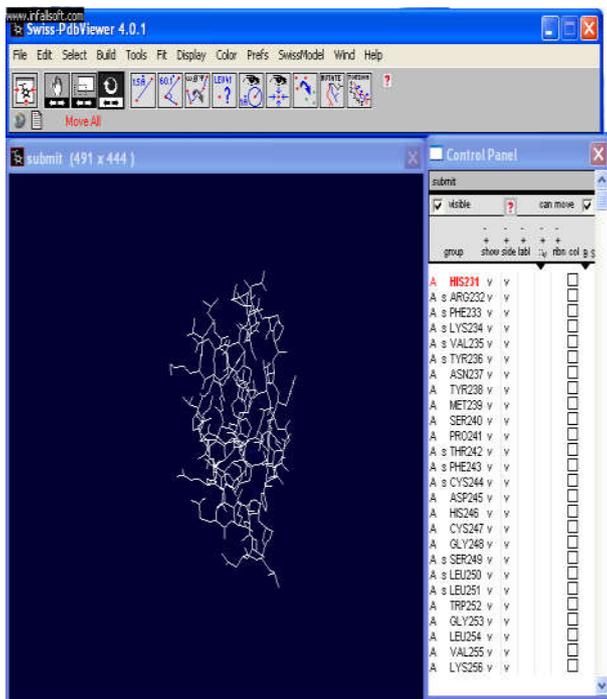
Residues in most favoured regions [A,B,L]	41	95.3%
Residues in additional allowed regions [a,b,l,p]	2	4.7%
Residues in generously allowed regions [-a,-b,-l,-p]	0	0.0%
Residues in disallowed regions	0	0.0%

Number of non-glycine and non-proline residues	43	100.0%
Number of end-residues (excl. Gly and Pro)	5	
Number of glycine residues (shown as triangles)	4	
Number of proline residues	1	

Total number of residues	53	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

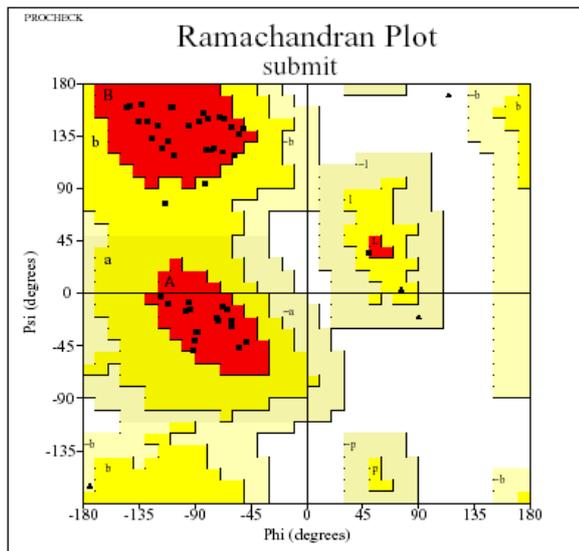
Structure of the modelled target protein using swisspdb viewer



Modelled protein in molgero molecular viewer



Accuracy of the modelled protein using procheck-ramachdran plot



RESULT OF MUSTER

MUSTER (MULTI-Source ThreadER) results

> RASSF1_340 residues, 'Hard' target

Your submitted sequence and predicted secondary structure (by PSI-PRED):

```

Seq: MSGEPELIELRELPAGRAGKGRTRLERANALRIARGTACNPTRQLVPGRGHRFPAGPFA
SS: CCCCCHHHHHHHHCCCCCCCCCCCCCHHHHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC
Pos: 123456789012345678901234567890123456789012345678901234567890
    
```

```

Seq: THTWCDLQDFIWGVRKGLQCAHCKFTCHYRCRALVCLDCGPRDLGWPEAVERDTNVD
SS: CCCCCCHHHHHHHHHHCCCEEECCCCCEEEHHHHHHCCCCCCCCCCCCCCCCCCCCCCCC
Pos: 123456789012345678901234567890123456789012345678901234567890
    
```

```

Seq: EPVWEWTPDLSQAEIEQKIKEYNAQINSNLFMSLNKDGYSYTGFIKVQLKLRPVSVPSSK
SS: CCCCCCCCCCCCCCHHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
Pos: 123456789012345678901234567890123456789012345678901234567890
    
```

```

Seq: KPPSLQDARRGPGRGTSVRRRTSFYLPKDAVKHLHLVLSRTRAREVIEALLRKFLVVDVDP
SS: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHHHHHHCCCCCE
Pos: 123456789012345678901234567890123456789012345678901234567890
    
```

```

Seq: KFALFERAERHQGVYLRKLLDDEQPLRLRLLAGPSDKALSFVLKENDSGEVNWDADFSPME
SS: HEEEEEEEECCCCCEEECCCCCCCCCHHHHHHHCCCCCCCCCCCCCCCCCCCCCHHHHHHH
Pos: 123456789012345678901234567890123456789012345678901234567890
    
```

```

Seq: LHNFLRILQREEEHRLRQLQKYSYCRQKIQEPALHACPLG
SS: HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCCC
Pos: 12345678901234567890123456789012345678901234567890
    
```

For detailed information about output, please see [readme file](#).

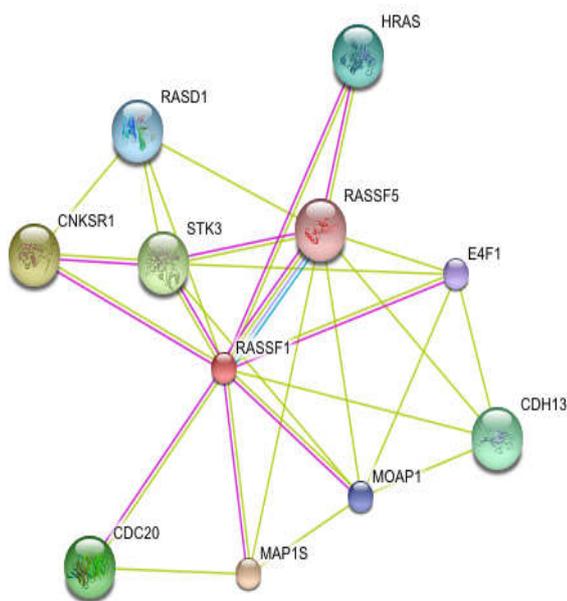
Models generated by MUSTER

MUSTER is a Multi-Source ThreadER program, which considers six different sources: (1) sequence-derived profiles; (2) secondary structures; (3) structure-derived profiles; (4) solvent accessibility; (5) torsion angles (psi and phi angles); (6) hydrophobic scoring matrix. The optimized threading is found by global dynamic programming.

Rank	Template	Align_length	Coverage	Zscore	Seq_id	Type	Target-template alignments	3-D models from threading alignments	Full-length models by MODELLER
1	1xa6A	296	0.870	4.513	0.159	Bad	alignment_1	threading_1	model_1
2	1f8ra	313	0.920	4.106	0.086	Bad	alignment_2	threading_2	model_2
3	2bf4A	315	0.926	4.093	0.098	Bad	alignment_3	threading_3	model_3
4	1xwl_	323	0.95	4.013	0.071	Bad	alignment_4	threading_4	model_4
5	1xpqA	332	0.976	4.006	0.120	Bad	alignment_5	threading_5	model_5
6	1foha	336	0.988	3.929	0.104	Bad	alignment_6	threading_6	model_6
7	1kfsa	321	0.944	3.919	0.084	Bad	alignment_7	threading_7	model_7
8	1t2A	327	0.961	3.841	0.083	Bad	alignment_8	threading_8	model_8
9	1f0ka	323	0.95	3.806	0.087	Bad	alignment_9	threading_9	model_9
10	1gosa	331	0.973	3.782	0.109	Bad	alignment_10	threading_10	model_10

Reference: S. Wu, Y. Zhang. MUSTER: Improving protein sequence profile-profile alignments by using multiple sources structure information. *Proteins: Structure, Function, and Bioinformatics* 2008; 72: 547-556

Target proteins interaction network



DISCUSSION

The model structure for RASSF1 protein using swissPDB Viewer and its analysis of ramachandran plot using SAVSERVER predicted that it is a fairly, very good model. RASSF1 protein is found to interact with the complex of CNK1, MST, MAP-1 in order to react with Bax protein to induce apoptosis which is Important for tumor suppression in Non Small Cell Lung Cancer. RASSF1 also causes cell cycle arrest by preventing the formation of complex between Cdc20 and APC thereby results in tumor suppression. The loss of

these control and deactivation causes Non Small Cell Lung cancer. A new therapeutic approach and path of activation of this protein in cellular function can be investigated based on this structure to prevent Non Small Cell Lung cancer or to reduce the cell tumourogenesis. The active site prediction and analysis of this protein can predict new modes to increase the interaction between this protein and other proteins, thereby gain resistance against cancer

REFERENCES

- Aldington S. *et al.* 2008. Cannabis use and risk of lung cancer: A case-control study. *European Respiratory Journal*, 31(2): 280-286.
- American Cancer Society, 2012. Cancer Facts and Figures 2012. Atlanta: American Cancer Society. Available online: <http://www.cancer.org/Research/CancerFactsFigures/CancerFactsFigures/cancer-facts-figures-2012>.
- Dasgupta P, *et al.* 2006. Nicotine inhibits apoptosis induced by chemotherapeutic drugs by up-regulating XIAP and survivin. *Proceedings of the National Academy of Sciences*, 103(16): 6332-6337.
- Ettinger DS. 2008. Lung cancer and other pulmonary neoplasms. In L Goldman, D Ausiello, eds., *Cecil Medicine*, 23rd ed., pp. 1456-1465. Philadelphia: Saunders Elsevier.
- Fox JL, *et al.* 2004. The effect of smoking status on survival following radiation therapy for non-small cell lung cancer. *Lung Cancer*, 44(3): 287-293.
- National Cancer Institute (2011). Small Cell Lung Cancer PDQ: Treatment-Patient Version. Available online: <http://www.cancer.gov/cancertopics/pdq/treatment/small-cell-lung/Patient>.
- Silverstri GA. and Jett JR. 2010. Clinical aspects of lung cancer. In R Mason *et al.*, eds., *Murray and Nadel's Textbook of Respiratory Medicine*, 5th ed., vol. 2, pp. 1116-1144. Philadelphia: Saunders.
- Temel JS. *et al.* 2010. Early palliative care for patients with metastatic non-small-cell lung cancer. *New England Journal of Medicine*, 363(8): 733-742.
- U.S. Department of Health and Human Services, 2010. A Report of the Surgeon General: How Tobacco Smoke Causes Disease-The Biology and Behavioral Basis for Smoking-Attributable Disease Fact Sheet. Available online: <http://www.surgeongeneral.gov/library/tobacco-smoke/factsheet.html>.
