In-silico study of small cell lung cancer based on protein structure and function: A new approach to mimic biological system

Nidhi Sood, Sameer Chaudhary,
Tanvee Pardeshi, Shama Mujawar,
Krishna Balaji Deshmukh,
Saba Afrin Sheikh, Preety Sharma

Department of Computational Chemistry,
RASA Life Science Informatics, Pune,
Maharashtra, India

Abstract

Lung cancer being the most common disease worldwide that leads to a number of deaths. A huge amount of effort has been done in screening trials for early diagnose treatment which increases the disease-free survival rate. Based on the expression of protein of mouse double minute 2 and tumor protein 53 complex, we have identified the antagonist for this complex that would facilitate the treatment for specific lung cancer. It is a complex disease that involves vast investigation for the characterization of a lung cancer and thus, computational study is being developed to mimic the *in vivo* system. In this work, a computational process was employed for the identification of these proteins, with a short and simple method to discover protein-protein interactions. Moreover, these proteins have more similarities in their function with the known cancer proteins as compared to those identified from the protein expression specific profiles. A new method that utilizes experimental information to improve the extent of numerical calculations based on free energy profiles from molecular dynamics simulation. The experimental information guides the simulation along relevant pathways and decreases overall computational time. This method introduces umbrella sampling simulations. A new technique umbrella sampling is described where the high efficacy100 of this technique enables uniform sampling with several degrees of freedom. Here, we review the protein interactions techniques and we focus on main concepts in the molecular of *in-silico* study in lung cancer. This study recruiting new methods proved the efficiency and showed good results.

Key words: Mouse double minute 2 and p53 complex protein, small cell lung cancer, umbrella sampling simulation

INTRODUCTION

Lung cancer is a malignant tumor characterized by uncontrolled growth of cells in tissues. As reported by

Address for correspondence:
Miss. Nidhi Sood,
RASA Life Science Informatics, 301,
Dhanashree Apartments, Opposite Chitaranjan Vatika,
Pune - 411 016, Maharashtra, India.
E-mail: nidhi29sood@gmail.com

WHO, with approximately 14 million new cases and 8.2 million cancer-related deaths observed in 2012. The number of new cases is expected to rise by about 70% over the next two decades. Lung cancer being the leading cause of cancer death and the second most common cancer among both men and women in the United States.[1]

Symptom for lung cancer detection often occurs at late stage when the prognosis is poor.[2,3] The major cause is smoking with the highest 85% of all cases of lung cancer. Harmful chemicals that are present in cigarette smoke are proven cancer inducer. Other factors may include radiations, asbestos etc., nearby pollution living areas as well as human immunodeficiency virus. About 10–15% of the lung cancers are small cell lung cancer (SCLC) unlike nonsmall cell lung cancer (NSCLC). The overall 5-year survival rate for NSCLC ranges from 9% to 15%.[4]

Tumor protein 53 (TP53) encoded by the tp53 gene. This protein is crucial in multicellular organisms, where it

regulates the cell cycle and thus functions as a tumor suppressor, preventing cancer. It regulates cell division by inhibiting them to divide in an uncontrolled manner. Mouse double minute 2 (MDM2) also known as E3 ubiquitin-protein ligase. Mdm2 is a protein encoded by the MDM2 gene and is an important negative regulator of the p53 tumor suppressor.[5] Mdm2 is phosphorylated at multiple sites in cells. Following DNA damage, phosphorylation of Mdm2 leads to changes in protein function and stabilization of p53.[6]

Phosphatase and tensin homolog (PTEN) is a protein encoded by the PTEN gene. PTEN act as a tumor suppressor gene through the action of its lipid phosphatase protein activity.[7] This phosphatase is involved in the regulation of the cell cycle, preventing cells from growing and dividing too rapidly. PTEN tumor suppressor protein inhibits activation of Akt, and this restricts Mdm2 to the cytoplasm.[8] The protein encoded by this gene is a phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase. It contains a tensin-like domain as well as a catalytic domain similar to that of the dual specificity protein tyrosine phosphatases. Similarly, Nutlin-3 is a small molecule inhibitor of the MDM2/p53 interaction.[9]

The current issues with computational biology have been prioritized by a large increase in the number of potential therapeutic targets willing to comply with an investigation. Thus, for pharmaceutical industries, they reveal new discoveries with highly efficient manner. There is an urgent need, therefore, to review the technologies currently employed in lead identification and critically assess methodologies which are likely to increase productivity at the early discovery stages. High-throughput screening has traditionally been most widely used methodology in the drug discovery process

The goal of this study was to determine the regulation of active genes or proteins with their normal functions by studying the mutated form as well of these proteins. We sought to address the upregulation of tumor suppressor proteins for the treatment of lung cancer. Therefore, we are aiming to achieve this with the study of *in-silico* methods that employ scenarios to look out for better options.

MATERIALS AND METHODS

Extensive literature and text mining were carried out to study lung cancer unambiguous protein inhibitors. The expected behavior of tp53 and mdm2 complex was responsible for lung cancer. The structures were available from Protein Data Bank (PDB). The information regarding these proteins were available in NCBI, UNI PROT. These proteins were then best-viewed under Discovery studio where water and heterogenous atoms were removed, followed by protein-protein docking and simulation.

For mdm2 and p53 protein complex with PDB ID 4HFZ, antagonist proteins were PTEN and Nutlin3 with PDB ID 1D5R and 4HG7, respectively.

Online servers
With the in-depth study of literature, online servers were used for protein-protein interaction or docking. Services were provided by online servers but with the best results, Hex, Clus–Pro, and Fire Dock servers were used.

Clus Pro server
Here, we described an automated rigid body docking which instantly filters docked conformations based on the protein parameters and introduced them according to their clustering properties. Filtered conformations involved the utilization of evaluation methods based on empirical free energy which selects the combination with lowest de-solvation and electrostatic energies. Clus Pro server available on http://cluspro.bu.edu/home.php.

Docked conformations have been generated using the docking program DOT based on the fast-Fourier transform (FFT) correlation approach. We have used version 1.0 alpha of DOT with a 45° angle increment, and default values of 1Å grid-step and 4Å surface layer.[10]

Fire dock server
Fire dock method involves scoring and flexible refinement for protein-protein docking solutions where the atomic contact energy and repulsive vanderwaals energy were of main concern. It includes a side-chain optimization component. It allows a high-throughput refinement of up to 1000 solution candidates. The method simultaneously targets the problem of flexibility and scoring of solutions produced by fast rigid-body docking algorithm. This server is available on http://bioinfo3d.cs.tau.ac.il/FireDock/.

Hex server
Hex server is the first FFT-based protein docking. It is easy to use and can upload protein structures in PDB format. For a blind unconstrained 6D docking run, it is recommended to use default values for all parameters.

Low energy solutions were clustered and identified the allocated entries for distinct orientations. The remaining solutions were then rescanned for clustering until all solutions have been grouped. These clustering parameters help to reduce the number of false-positives generated docking search. All calculations use the same parameters except for the ligand cut off angle which varies. It recognizes known complexes and then helps in reconstructing these known complexes.

Umbrella sampling
Molecular dynamics (MD) simulations provide detailed information for a protein of known structure. It employed

the technique to calculate the three dimensional structure of a complex which determine which proteins interact. There are several tools for protein-protein docking, one of which is umbrella sampling which uses critical assessment of predicted interactions (CAPRI) blind docking method. With the applications of Gromacs, umbrella sampling version 4.5.6 used to determine the conformations. It is normally obtained by weighted histogram analysis method (WHAM). It enables the uniform sampling for MD with the conformational space and several degrees of freedom.

The WHAM provides the potential mean force with accurate statistics as well as efficient utilization for additional simulations to reduce errors. This method advances significantly the execution of recombining the various windows in complex arrangement.

RESULTS AND DISCUSSION

In this section, we presented results for free energy umbrella sampling technique. The purpose of this study was to validate the conformations determined from umbrella sampling. The protein complex formed based on Ramchandran plot was further encouraged to perform simulation using Gromacs.

An MD simulation was performed on a protein complex which showed different potentials of mean force. In this section, the results for free energy landscapes of mdm2 and p53 protein complex to assess the performance of this method were presented.
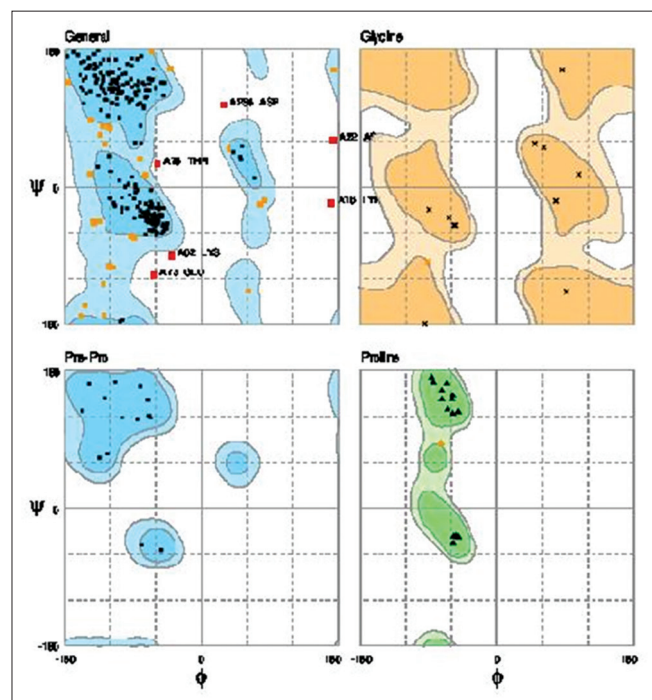
Figure 1: 1D5R protein structure on the basis of Ramchandran plot where number of residues in favored, allowed, and disallowed regions were described by, where it validates protein structures

Primary structural analysis

On the basis of PDB structure, protein 4HFZ, that is, complex protein of mdm2 and p53 was selected. As shown in the Figure 1, the Ramchandran plot shows the phi-psi torsion angles for all residues in the structure. According to Ramchandran plot, number of residues in favored region should be more than 90%. For the number of residues in allowed region should be 90%. The darkest areas correspond to the "core" regions representing the most favorable combinations of phi-psi values. Ideally, one would hope to have over 90% of the residues in these "core" regions, and it is one of the better guides to stereochemical quality.[11]

Figure 1 illustrates the selection of proteins based on Ramchandran's plot. It provides an overview of allowed and disallowed regions of torsion angle values, serving the importance for the quality of protein three dimensional structures. Basically, it determines the compatibility of an atomic model three-dimensional structure based on its location and environment (alpha, beta, polar etc.,) and comparing the results.

Similarly, Figure 2 shows the Ramchandran plot assessment (RAMPAGE) analysis for 1D5R and 4HG7. The final structures were obtained by taking the best structures complex with each other on Hex, Clus Pro and Firedock servers. For Hex servers, two proteins 1D5R and 4HG7 were recruited to bind where 15 possible best structures were provided, and Hex calculates an excluded volume model of shape complementarity with an optional *in vacuo* electrostatic contribution.[12] For Clus Pro server, protein 1d5r and 4hg7 were subjected to their binding sites and 10 best structures were found to be most best studied. The summary of results obtained by Clus Pro server 2.0, the quality of results evaluated by CAPRI. Best 10 structures were uploaded on the basis of binding efficiency.

Figure 2 shows that the protein has passed the result with 98.37% of the residues. Similarly, for Figures 3 and 4 show the RAMPAGE analysis of protein 4HG7 and 4HFZ. These both proteins have concluded that with both the
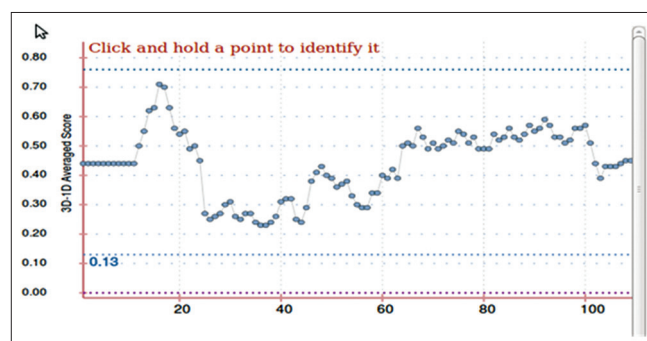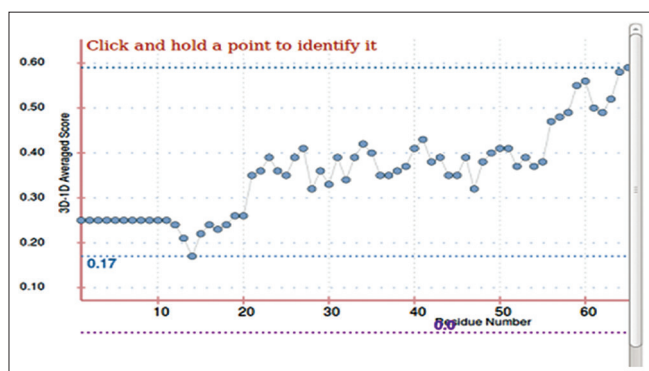

Figure 2: 1D5R Ramchandran plot assessment where 98.37% of the residues had an average 3D-1D score ≥0.2 and at least 80% of the amino acids have scored ≥0.2 in the 3D/1D profile

experimental and theoretical data interpretation, these proteins provide useful preclinical results with the relevance of further development work related to various treatments.

Analysis of servers is given in Table 1 where the least energy docking solutions were selected. From the docked solutions, out of all complexes, top 15 best results were chosen for further simulation processes. Clus Pro analysis also provided the best-docked conformations where out of 300, best 9 were taken and analyzed on their lowest energy basis.

Firedock server includes Table 1 of all the input solutions with PDB structures 4HG7 and 4 HFZ. The table is sorted according to global energy values, where 100 lowest energy structures were generated. With the application of Jmol, these structures can be viewed and downloaded as PDB files. This table provides the information of linear combination of normal modes for the receptor and ligand that produces the refined backbone conformation.

On the given results for these servers, The Hex server was best from all of them. As the results obtained in a specific manner that tabulates theoretical information sufficient enough
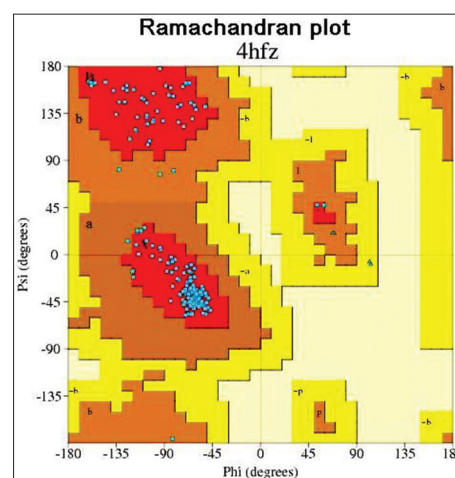
required for results to interpret. It selects the minimum energy pose and analyzes the docking results in Table 2.

After the docking of predicted protein files, MD simulation was performed based on the absolute quality of the models obtained to refine them for simulations. The final structures were produced by running the simulation program. The postsimulated structures were compared with presimulated structures for their validation. As from previous results for Ramchandran plot analysis, unfavorable residues from proteins were removed during MD simulations. With the efficient process for umbrella sampling, many files have been generated through which results have been predicted.

Docking analysis

Here, we analyzed data from a set of MD simulations. We construct histograms in Figure 5 with a uniform width and trajectories as the input for wham calculation. Due

Figure 3: 86.81% of the residues had an averaged3D/1D score ≥0.2 and at least 80% of the amino acids have scored ≥0.2 in the 3D/1D profile

Figure 4: Illustrates the Ramchandran plot for protein analysis for 4 HFZ. The dark blue, dark orange, dark blue for pro-pro and dark green color represents the general favored region and light color for general allowed regions respectively for general, glycine, pro-pro and proline residues

Table 1: The refined structures for fire dock server based on their binding energy

		Fire dock					
		Receptor		**Ligand**			
		New4HFZ.pdb		**New4HG7.pdb**			
Rank	**Solution number**	**Global energy**	**Attractive VDW**	**Repulsive VDW**	**ACE**	**HB**	**Structure**
1	7	−16.49	−42.53	39.34	8.03	−4.98	✓
2	6	−15.04	−27.17	14.33	7.99	−0.86	
3	8	−1.51	−24.93	12.25	18.02	−3.25	
4	3	6.76	−19.53	15.51	6.23	−3.30	
5	10	7.48	−20.47	5.29	10.93	−0.24	
6	5	15.87	−14.38	8.92	8.73	−0.24	
7	9	19.85	−29.88	40.11	16.44	−5.27	
8	1	24.21	−1.95	0.00	2.43	0.00	
9	4	27.37	−0.44	0.00	1.49	0.00	
10	2	34.90	−25.00	40.66	7.48	−3.65	

Table 2: The binding energies for Hex, Clus Pro, and Fire Dock server

Binding energy	Hex server	Clus Pro server	Fire Dock server
E total	−362.4	−642.4	−16.49
E total	−360.36	−672.6	−15.04
E total	−359.38	−658.9	−1.51
E total	−354.22	−615.5	6.76
E total	−348.75	−678.0	7.48
E total	−346.96	−609.6	15.87
E total	−341.15	−647.6	19.85
E total	−340.58	−618.5	24.21
E total	−339.19	−607.9	27.37
E total	−332.02	−609.0	34.90

Best top 10 refined structures where hex server was selected based on its binding efficiency
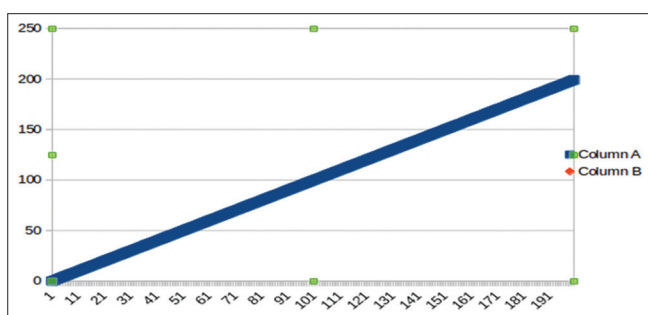
Figure 5: Histogram graph for count versus z axis which shows the protein binding

to significant variation as such reaction coordinates were chosen. The weighted histogram analysis method was used to calculate the PMF. The short trajectories were confined to explore the vicinity of a frontier point by means of a harmonic bias potential based on umbrella sampling data.

CONCLUSION

In-silico experimentation modeling of lung cancer involved predictions from biological data with computer-based models to mimic biological system to have investigations based on entirely computer methods. In this paper, we have provided the concept of *in-silico* study and its importance in the field of providing the structures to enhance computational methodology. *In-silico* study deals which is relevant to study SCLC and its applications with experimental research. This computational tool has the vast range to allow researchers to refine their experimental data with reducing costs and time, and increasing the research efficacy.

In this study, we introduced a computational method based on protein-protein interactions where we identified cancer-causing proteins. We applied this method to detect and study the SCLC and can find the best possible path to determine the consequences. Analysis of these proteins was

carried out and thence, their expressions were found to be involved in SCLC. In this analysis, bioinformatics approach have been identified that may be effective in developing protein interactions.

In the present work, as we have taken the proteins that play a vital role in SCLC and the proteins that were used against lung cancer to depict its efficacy by studying its interacting targets. From the results reviewed, we can conclude that rather than to opting for commercial treatments, these modified proteins were best-proven. For their effectiveness, these proteins can be further validated for clinical trials. This study enhances the initiation of protein simulations for lung cancer for future research work that provide social benefits. In addition, the results with Hex server as well as Clus Pro were of also greater reliability. This signifies the server to perform a full docking method in a fully automated manner.

REFERENCES

1. Kanagaraj B, Prakash A, Wadhwa G. An insight to virtual ligannd screening methods for structure based-drug design and methods to predict protein structure and function in lung cancer: Approaches and progress. J Crit Rev 2014;1:10-24.

2. William WN Jr, Lin HY, Lee JJ, Lippman SM, Roth JA, Kim ES. Revisiting stage IIIB and IV non-small cell lung cancer: Analysis of the surveillance, epidemiology, and end results data. Chest 2009;136:701-9.

3. Rami-Porta R, Crowley JJ, Goldstraw P. The revised TNM staging system for lung cancer. Ann Thorac Cardiovasc Surg 2009;15:4-9.

4. Petersen I. The morphological and molecular diagnosis of lung cancer. Dtsch Arztebl Int 2011;108:525-31.

5. Ryan KM, Phillips AC, Vousden KH. Regulation and function of the p53 tumor suppressor protein. Curr Opin Cell Biol 2001;13:332-7.

6. Shieh SY, Ikeda M, Taya Y, Prives C. DNA damage-induced phosphorylation of p53 alleviates inhibition by MDM2. Cell 1997;91:325-34.

7. Myers MP, Pass I, Batty IH, Van der Kaay J, Stolarov JP, Hemmings BA, *et al.* The lipid phosphatase activity of PTEN is critical for its tumor supressor function. Proc Natl Acad Sci U S A 1998;95:13513-8.

8. Cantley LC, Neel BG. New insights into tumor suppression: PTEN suppresses tumor formation by restraining the phosphoinositide 3-kinase/AKT pathway. Proc Natl Acad Sci U S A 1999;96:4240-5.

9. Mashiach E, Schneidman Duhovny D, Andrusier N, Nussinov R, Wolfson HJ. FireDock: A web server for fast interaction refinement in molecular docking. Nucleic Acids Res 2008;1:36(Web Server issue): W229 32. doi: 10.1093/nar/gkn186. Epub 2008 Apr 19. PMC2447790.

10. Shangary S, Wang S. Small-molecule inhibitors of the MDM2-p53 protein-protein interaction to reactivate p53 function: A novel approach for cancer therapy. Annu Rev Pharmacol Toxicol 2009;49:223-41.

11. Morris AL, MacArthur MW, Hutchinson EG, Thornton JM. Stereochemical quality of protein structure coordinates. Proteins 1992;12:345-64.

12. Ritchie DW, Kemp GJ. Protein docking using spherical polar Fourier correlations. Proteins 2000;39:178-94.