Protein Modeling and Sequence Analysis Studies on Human Lung Cancer Potential Protein Target (Crp C-Reactive Protein, Pentraxin-Related) Using Bioinformatics Protocols

Vasiya thanveer1 and Balaji.M2

1Department of Biotechnology, DKM college for women, vellore . 
2Director , Akshaya Neuroinformatics Research center Pvt ltd., Chennai.

INTRODUCTION
Worldwide, lung cancer is the most common cancer in terms of both incidence and mortality (1.1 million new cases per year and 0.95 million deaths in males and 0.51 million new cases per year and 0.43 million deaths in females). The highest rates are in Europe and North America1. The population segment most likely to develop lung cancer is over-fifties who have a history of smoking. Lung cancer is the second most commonly occurring form of cancer in most Western countries, and it is the leading cancer-related cause of death. In contrast to the mortality rate in men, which began declining more than 20 years ago, women's lung cancer mortality rates have been rising over the last decades, and are just recently beginning to stabilize. The evolution of “Big Tobacco” plays a significant role in the smoking culture. In this research study we found out the potential mutation gene involved in Human Lung cancer disease and to perform protein sequence analysis, protein modelling and ligand binding sites prediction using Bioinformatics tools. There are three steps involved in our present research investigation. The protein primary structure analysis, protein three dimensional structure prediction and drug binding sites identification. The identified drug binding sites of the (CRP C- REACTIVEPROTEIN, PENTRAOXIN-RELATED) protein are the best potential inhibitors for drug docking studies. The identification of drug (ligand) binding sites is an important step in insilico structure based drug designing and in the field of cheminformatics. The targets discovered by us would certainly prove to be useful in treating Lung cancer disease in Human beings.

Abstract
Lung cancer is the second most commonly occurring form of cancer in most Western countries, and it is the leading cancer-related cause of death. In contrast to the mortality rate in men, which began declining more than 20 years ago, women’s lung cancer mortality rates have been rising over the last decades, and are just recently beginning to stabilize. The evolution of "Big Tobacco" plays a significant role in the smoking culture. In this research study we found out the potential mutation gene involved in Human Lung cancer disease and to perform protein sequence analysis, protein modelling and ligand binding sites prediction using Bioinformatics tools. There are three steps involved in our present research investigation. The protein primary structure analysis, protein three dimensional structure prediction and drug binding sites identification. The identified drug binding sites of the (CRP C- REACTIVEPROTEIN, PENTRAOXIN-RELATED) protein are the best potential inhibitors for drug docking studies. The identification of drug (ligand) binding sites is an important step in insilico structure based drug designing and in the field of cheminformatics. The targets discovered by us would certainly prove to be useful in treating Lung cancer disease in Human beings.

Materials and Methods
NCBI-PMC, NCBI – Pubmed, NCBI – OMIM, Protparm, CPH 3.0 model server, Discovery studio software, Molsoft, 3D LIGAND SITE

Literature collection:
The molecular details (literature) of Lung cancer protein (CRP C- REACTIVEPROTEIN, PENTRAOXIN-RELATED) were collected using online digital libraries and literature databases like OMIM, NCBI – PUBMED and PUBMED Central database.

Sequence retrieval system:
The protein sequence of CRP protein sequence was retrieved from the NCBI FASTA format-protein, in order to perform the protein sequence analysis and modeling.

▼ To whom correspondence should be addressed: Vasiya thanveer Email: bioinfobalaji@gmail.com
Primary sequence analysis:
The primary sequence analysis of CRP protein was determined using EXPASY – PROTPARM tool in order to identify the physiochemical properties of the input protein sequence.

Peptide 3Dimensional structure prediction:
The PHA – motif peptide sequences were converted into 3D structure using protein homology modeling software - CPH 3.0 model servers.

Molecular visualization tool:
The predicted 3 D structure was viewed with help of molecular visualization tools like Discovery studio and molsoft.

Protein structure validation:
The predicted modeled peptide structures were validated using Assessment of Ramachandran plot server – RAPPER.

Protein –ligand binding sites prediction:
The modeled 3d protein structure was applied into 3dligndsite server in order to identify the molecular binding sites present in the CRP protein structure.

<p>| Table: 1 |</p>
<table>
<thead>
<tr>
<th>Gene name</th>
<th>OMIM ID</th>
<th>Gene ID</th>
<th>CHR LOC</th>
<th>PDB</th>
<th>Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP C- REACTIVE PROTEIN, PENTRAXIN-RELATED</td>
<td>123260</td>
<td>1401</td>
<td>1</td>
<td>no</td>
<td>The protein encoded by this gene belongs to the pentaxin family. It is involved in several host defense related functions based on its ability to recognize foreign pathogens and damaged cells of the host and to initiate their elimination by interacting with humoral and cellular effector systems in the blood. Consequently, the level of this protein in plasma increases greatly during acute phase response to tissue injury, infection, or other inflammatory stimuli. [provided by RefSeq, Sep 2009]</td>
</tr>
<tr>
<td>PTX1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure: 1Protein FASTA sequence

>gi|17530543|gb|AAL40835.1|AF449713_1 C-reactive protein, pentraxin-related [Homo sapiens] MEKLLCLFLVLTSLSHAFGQTDMRSRKFVEFPKESDTSVYSLKAPLTKPLKAFVTCLHYFTLSSRGRYSIF SYATKROQDNEILIFWSKDIGYFTVGGSELFEVPEVTAPVHICTSWESAGIVEFVWDGKPRVRKSLK KGVTVGEAASILGQEQDSFGGNGSFQSLVGDIGNVNMWDFVLSDEINTIYLGPGFSPANVNLNWRALKYEVQGEVFTKPQLWP
Figure: 1.1 The primary structure analysis PROTPARM tool

Theoretical pI: 5.45

Amino acid composition:
- Ala (A) 10 4.5%
- Arg (R) 6 2.7%
- Asn (N) 7 3.1%
- Asp (D) 9 4.0%
- Cys (C) 3 1.3%
- Gin (Q) 7 3.1%
- Glu (E) 15 6.7%
- Gly (G) 18 8.0%
- His (H) 3 1.3%
- Ile (I) 12 5.4%
- Leu (L) 20 8.9%
- Lys (K) 14 6.2%
- Met (M) 3 1.3%
- Phe (F) 16 7.1%
- Pro (P) 11 4.9%
- Ser (S) 23 10.3%
- Thr (T) 18 6.2%
- Trp (W) 6 2.7%
- Tyr (Y) 8 3.6%
- Val (V) 19 8.5%
- Pyl (O) 0 0.0%
- Sec (U) 0 0.0%

Total number of negative: 0.0%
Total number of positive: 0.0%

Total number of negatively charged residues (Asp + Glu): 24
Total number of positively charged residues (Arg + Lys): 20

Atomic composition:
- Carbon (C): 1181
- Hydrogen (H): 1746
- Nitrogen (N): 202
- Oxygen (O): 332
- Sulfur (S): 6

Formula: C₁₁₈₁H₁₇₄₆N₂₀₂O₃₃₂S₆
Total number of atoms: 3817

Instability index:
The instability index (II) is computed to be 97.34.
This classifies the protein as stable.

Aliphatic index: 94.78
Grand average of hydropathicity (GRAVY): -0.332
Figure: 1.2 PDB structure database

Figure: 2 Protein Modeling server – 3 Dimensional structure of (CRP C-REACTIVEPROTEIN, PENTRAXIN-RELATED) protein structure prediction
Template selection and modeling

Figure: 3 Dimensional structure of (CRP C- REACTIVEPROTEIN, PENTRAXIN-RELATED) protein

Discovery studio software

CPK model – color atoms: Red – oxygen, Blue – nitrogen, Grey – Carbon and Yellow – sulphur
The N termini of proteins are colored blue (similar to the CPK color for nitrogen) and the C termini, red (similar to the CPK color for oxygen).

hydrophobic    yellow1  ala,val,phe,pro,met,ile,leu
polar pink     ser,thr,tyr,his,cys,cyss,asn,gln,try,glu
charged (+)    blue lys,arg
charged(-)     red  asp,glu
Conclusion

There are three steps involved in our present research investigation. The protein primary structure analysis, protein three dimensional structure prediction and drug binding sites identification. The identified drug binding sites of the (CRP C- REACTIVEPROTEIN, PENTRAXIN-RELATED) protein are the best potential inhibitors for drug docking studies.
The identification of drug (ligand) binding sites is an important step in in silico structure-based drug designing and in the field of Cheminformatics. This work would be useful in the field of Pharmacophore designing and Pharmacinformatics studies. The targets discovered by us would certainly prove to be useful in treating Lung cancer disease in Human beings.

References