



A Computer Aided Drug Designing Approach to Target the KRAS protein with Ligand DON to treat Lung Cancer

*Brwa Abdulrahman Abubaker

Yüzüncü Yıl Üniversitesi, Turkey

Email: brwa.pshdary@gmail.com, Contact: 00964-7504035635

Received: 5th June 2015, Accepted: 20th June 2015, Published: 1st September 2015

Abstract:

Development of Drugs towards ailments was one of the oldest techniques followed since ages. Though the technique was proved to be effective one of the major drawbacks is the time involved in development of a drug using traditional Methods. The Computer Aided Drug Designing or the Rational Drug designing is one of the best solutions for the problems faced by the traditional method. The current work aims to develop a potential drug molecule to target the K RAS protein, vital in the onset of cancer. This aim is achieved by employing the CADD approach. The sequence of the K RAS was retrieved from NCBI data base and the sequence analysis was performed using various bioinformatics tools to annotate the protein. The Functional study of the protein was also performed using SMART, ProtParam, Fingerprint Scan and CDD analysis. The Structural study was performed using ExPASy's tools. The secondary confirmations in the protein were studied using SOPMA and the tertiary structure's PDB Id was obtained using Phyre. The 3D structure of the protein was obtained from RCSB Protein Data Bank. The structure was visualized using RASMOL software. The Lead chemicals were collected from PubChem Data base. The screening of the chemicals was performed for various parameters like Drug Likelihood and Ambiguity. The final list of chemicals was processed for docking with the K RAS receptor using HEX 8.00 Software. Based on the docking result, the best drug selected for the KRAS receptor was DON with the CID 5359375.

Keywords:

CADD, SMART, DON, K Ras, Docking

Introduction:

KRAS, also called Kras, K-RAS, or KRAS2, is a gene that is important in cell growth and tumor development. The KRAS gene is mutated in about 25% of NSCLC cases. Patients are usually tested for the KRAS and EGFR biomarkers at the same time, because the presence of both (Complete statement), Approximately 15–25% of patients with lung adenocarcinoma have tumor associated *KRAS* mutations. *KRAS* mutations are uncommon in lung squamous cell carcinoma. Several approaches have been implemented in relation to the cancer therapy. The use of

bioinformatics in cancer therapy is emerging very drastically and has gained full attention.

In Computer aided drug designing or rational drug design, biologically active compounds are specifically designed or chosen to work with a particular drug target. This method often involves the use of molecular design software, which researchers use to create three-dimensional models of drugs and their biological targets [3]. Drug designing requires the Chemical leads to be screened which are then obtain from various databases and sources. One such database is the Pubchem. This is the product of NCB. This contains chemical molecules along with the annotations about their bioactivity, Molecular Properties, Analogues etc. All the chemical from this database are also linked to their uses and application till the date of their updation. Each chemical in Pubchem has been assigned a CID No which represents the structure.

Docking is a phenomena of a ligand binding to its receptor. Protein-protein interactions play a central role in various aspects of the structural and functional organization of the cell, signal transduction, and gene regulation. In Drug designing a dock is performed between a chemical lead and the protein receptor to test the binding feasibility and analyze the energy of the pair. Based on the docking results the best ligand can be selected, however the function of the ligand is a separate issue to be dealt with.

Materials and Methods:

Sequence Retrieval:

Sequence retrieval is a protocol for the selection and retrieval of a gene or a protein sequence from the required database. NCBI is the major master database used for this purpose. However other secondary databases like Uniprot, Swissprot etc also are available. In NCBI sequence retrieval provides information related to the source, Organism, Length, Type of the sequence etc. The data can be further used for the analysis.

SMART Domain Analysis:

It is a research tool used for the annotation of the sequence in terms of the regions of importance on it. It is preferably used for the identification of total domains and each of its function. (a Simple Modular Architecture Research Tool). More than

500 domain families found in signalling, extracellular and chromatin-associated proteins are detectable. Each domain found in a non-redundant protein database as well as search parameters and taxonomic information are stored in a relational database system. User interfaces to this database allow searches for proteins containing specific combinations of domains in defined taxa.

FingerPrint Scan for Family Identification:

This is an online tool, a product of EMBL used for the annotations of Fingerprints, a conserved patterns on the proteins. The identified fingerprints are further classified as per the family of protein and its functionality. The Fingerprint scan contains a database or its link for the protein domains and families to provide the user required output.

Physico Chemical Characterization using ProtParam :

ProtParam from ExPASy, is used for the complete physicochemical characterization of the proteins. This tool computes various physico-chemical properties that can be deduced from a protein sequence. The parameters computed by ProtParam include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).

SOPMA for Secondary Structural Conformation of the Protein:

Also a tool from ExPASy, performs significant Protein secondary structure prediction by consensus prediction from multiple alignments.

Phyre for Tertiary Structure Analysis:

Phyre2 is a suite of tools available on the web to predict and analyze protein structure, function and mutations. The focus of Phyre2 is to provide biologists with a simple and intuitive interface to state-of-the-art protein bioinformatics tools. Phyre2 replaces Phyre, the original version of the server for which we previously published a paper in Nature Protocols.

RASMOL for 3D Visualization:

It is a molecular graphics program for the visualisation of proteins, nucleic acids and small molecules. The program is aimed at display, teaching and generation of publication quality images. RasMol is a computer program written for molecular graphics visualization intended and used primarily for the depiction and exploration of biological macromolecule structures, such as those found in the Protein Data Bank.

Pubchem for Chemical Library Generation:

Unlike the others, Pubchem is a database or the search engine, an archive containing the all known

chemicals and substances along with their complete structural and functional annotation. PubChem's integration with NCBI's Entrez search engine provides sub/superstructure, similarity structure, bioactivity data, and other searching.

Lipinski rule for Screening of Drug likeliness:

The rule depicts the basic parameters that have to be exhibited by a chemical to make it positive for drug likeliness. All the four parameters of Lipinski, The Molecular weight, H bond donor, H bond acceptor and log P values must follow their threshold values. The cut off values are either 5 or multiples of five thus, the rule is Lipinski rule of 5.

Ambiguity screening using Dundee Pro Drg Server:

An online tool for the calculation of chemical ambiguity with respect to Bonds, Bond Angles, Partial Charges, Improper Dihedrals and Dihedrals. Ambiguity refers to the degree of non standard nature. Ambiguity value for an Ideal chemical should be Zero, however no chemical would be ideal, thus should possess the lowest ambiguity values. The little ambiguity existing in the chemical can be minimized by a process called Geometry Optimization.

Hex for Molecular Docking:

Hex is an interactive molecular graphics program for calculating and displaying feasible docking modes of pairs of protein and DNA molecules. Hex can also calculate protein ligand docking, assuming the ligand is rigid, and it can superpose pairs of molecules using only knowledge of their 3D shapes.

Results and Discussion:

Sequence Retrieval:

The protein sequence of K Ras has been retrieved from NCBI data base. The length of the sequence was found to be 189 amino acids with its accession number NP_203524. The total number of CDD regions identified was 13, out of which one was PTZ00369 and the other was H_N_K Ras Like with its 12 copies on the same gene.

SMART Domain Analysis:

After performing the SMART with the Ras sequence it was identified that the total number of domains to be only 1 i.e the RAS domain. The domain was found in the region 1 to 166.

Fig 1: Showing the RAS domain of K RAS protein



Physico Chemical Characterization using ProtParam :

The RAS sequence was submitted to the ProtParam and the results indicate that it is a Neutral protein with Polar Nature. It has the molecular wt of 21655.8, Iso electric point of 6.33 and the instability index of 38.58 making it a stable protein.

Secondary Structure Analysis by SOPMA:

The Sequence was analyzed to check for its secondary structural conformations. The details are mentioned in the Fig 2.

Fig 2: Secondary Structural Analysis- SOPMA:

Alpha helix	(Hh) :	81 is	42.86%
3 ₁₀ helix	(Gg) :	0 is	0.00%
Pi helix	(Ii) :	0 is	0.00%
Beta bridge	(Bb) :	0 is	0.00%
Extended strand	(Ee) :	39 is	20.63%
Beta turn	(Tt) :	19 is	10.05%
Bend region	(Ss) :	0 is	0.00%
Random coil	(Cc) :	50 is	26.46%
Ambiguous states (?)	:	0 is	0.00%
Other states	:	0 is	0.00%

The above fig 2 shows that the RAS has highest Alpha Helix confirmations with 81% followed by Random coil of 50%. The other confirmations include Extended strand of 38% and Beta turn with 19%.

Tertiary structure Prediction and Rasmol Visualization:

To retrieve the 3D structure of KRAS Phyre has been used. As per the results of Phyre the pdb id of the protein was found to be 1CTQA The structure was further visualized using RASMOL software.

Fig 3: 3D Structure of KRAS as visualized in RASMOL:



The above picture shows the 3D structure of the Protein in Rasmol

Chemical Analysis:

A total number of 120 chemicals were collected from the Pubchem database. All these were subjected to Primary screening based on Lipinski rule to detect their drug likeliness. All the chemicals selected in this step were further screened for Ambiguity using Dundee ProDrg Server. Finally a single chemical DON was selected which was used for docking with the selected receptor K RAS.

Docking of the ICTQ with the selected chemical DON in HEX SOFTWARE

Fig 4: The HEX window shoing the undocked Receptor and the Ligand

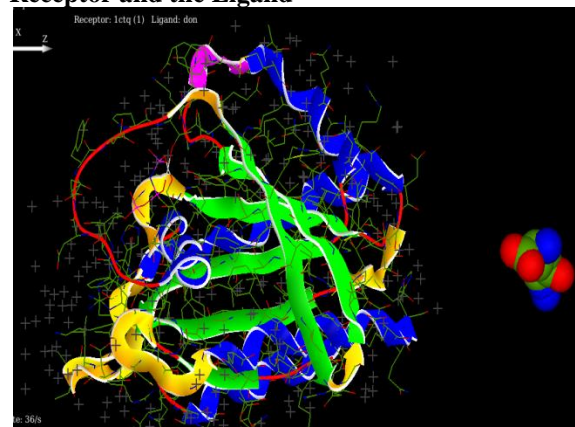
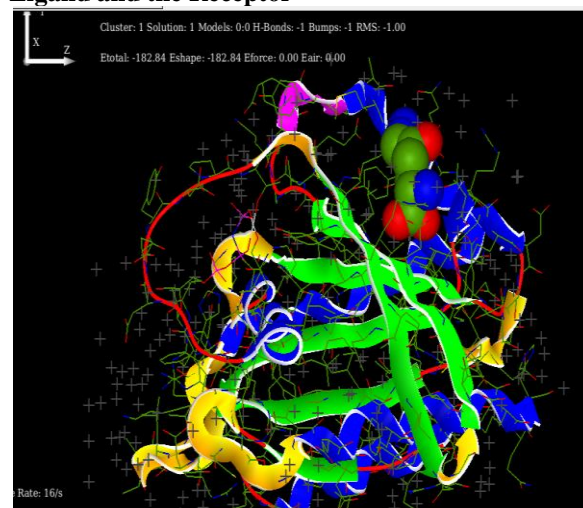


Fig 5: HEX software showing the Docked Ligand and the Receptor



Conclusion:

The current work aimed to study the KRAS protein ,one of the major causes of Lung Cancer. KRAS protein was retrieved from NCBI database. The sequence, Functional and Structural analysis was performed using various insilio tools and data bases. The 3D structure of the protein was downloaded from RCSB Protein databank whose PDB id was obtained from Phyre tool. The PDB id selected for the KRAS protein was 1CTQ having

95% similarity with the sequence. The structure was visualized using RASMOL software. The chemicals were obtained from the Pubchem data base. All the chemicals were screened based on the functionality. The selected chemicals were further screened for Drug likeliness based on Lipinski rule. All the final selected leads were screened for Ambiguity using Dundee ProdrG Server. The final chemical selected was DON with the CID no CID 5359375. Docking was performed between DON and KRAS in the software Hex. The docking energy was calculated to be **-182.84** which is an indication of a good docking couple. The current study concludes that DON can be used for targeting the KRAS protein involved in Lung cancer.

Acknowledgement:

The author would thank BioAxiD DNA Research Centre Pvt, Ltd, Hyderabad, India, for its help in providing all the resources in the successful completion of the project. The Author would also thank the Guide Mrs. Desu Jyothsna of BDRC for guiding in the complete research work.

References:

- 1) http://www.lungcancerprofiles.com/lung_cancer_and_biomarkers
- 2) <http://www.mycancergenome.org/content/disease/lung-cancer/kras>
- 3) <http://www.wisegeekhealth.com/what-is-rational-drug-design.htm>
- 4) https://en.wikipedia.org/wiki/Sequence_analysis
- 5) <http://smart.embl-heidelberg.de/help/latest.shtml>
- 6) <http://vlab.amrita.edu/?sub=3&brch=275&sim=1455&cnt=1>
- 7) <http://www.vivo.colostate.edu/hbooks/genetics/biotech/basics/prostruct.html>
- 8) <https://library.caltech.edu/learning/classhandouts/PubCheminfo.pdf>
- 9) https://en.wikipedia.org/wiki/Macromolecular_docking
- 10) <http://www.nature.com/nprot/journal/v10/n6/full/nprot.2015.053.html>
- 11) <https://en.wikipedia.org/wiki/RasMol>
- 12) <http://www.aacp.org/governance/SECTION/chemistry/Documents/Special%20Projects%20and%20Information/Chemistry%20Resources%20of%20Interest.pdf>
- 13) http://hex.loria.fr/manual800/hex_manual.pdf