Int J Pharm Bio Sci 2013 Jan; 4(1): (B) 370 - 382

Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

VIRTUAL SCREENING APPROACH OF DRUG DESIGNING FOR PARKINSON'S DISEASE

KULDEEP SAHU*[1] AND DR. SATPAL SINGH BISHT

Department of Biotechnology, Roland Institute of Pharmaceutical Sciences
Berhampur-760010, Orissa, India

ABSTRACT

Parkinson's disease *PARK1*, caused by mutations in the *SNCA* gene, which codes for the protein alpha-synuclein. *SNCA* gene is analysed by different NCBI tools and bioinformatics softwares (ORF, Map Viewer, e-PCR, Vec screen, Genscan, BLAST, FASTA, MSA, ClustalW, bioedit software and phylodraw software). Then protein information and protein structures are analysed by different proteomics tools and different softwares (Protparam, Protscale, GOR, SOPMA, SignalP, NetNGly, NetOGly, NetAcet, NetPhos, sulfinator, SOSUI, bioedit software, SPDBV software, RASMOL 3D analysis software). Homology modelling of the protein is done using SPDBV (SWISS PDB VIEWER). Active site analysis is done through Q-site finder method and the active site amino acids are noted. Standard available market drugs targeting the protein were identified as Carbidopa, Dopamine, Levodopa, Memantine, and Apomorphine. 16 similar molecules for these standard molecules were modelled using Argus Lab software. A database is created with all these molecules in Vega ZZ software. Virtual screening of these drugs is done through protein-database docking method. QSAR analysis is done through Hyperchem software. UV & IR transitions are done through CAche software.

KEY WORDS: ***SNCA* gene, *Alpha-synucein*, QSAR, Argus lab, SPDBV, Hyperchem and CAche.**

MR. KULDEEP SAHU
Department of Biotechnology, Roland Institute of Pharmaceutical Sciences
Berhampur-760010, Orissa, India

Corresponding author

INTRODUCTION

Parkinson's disease is a progressive disorder of the central nervous system. The motor symptoms of Parkinson's disease result from the death of dopamine-generating cells in the substantia nigra, (a region of the midbrain). It affects movement (motor symptoms). Typical other symptoms include disorders of mood, behaviour, thinking, and sensation (non-motor symptoms). There are four major dopamine pathways in the brain. The nigrostriatal pathway mediates movement and is the most conspicuously affected in early Parkinson's disease. The other pathways are the mesocortical, the mesolimbic, and the tuberoinfundibular. These pathways are associated with, respectively volition and emotional responsiveness, desire, initiative, and reward, and sensory processes and maternal behaviour. Neuropsychiatric pathology associated with Parkinson's disease (*PARK1)* caused by mutations in the *SNCA* gene, which codes for the protein alpha-synuclein. [1][2][3][4] Virtual screening as "automatically evaluating very large libraries of compounds" using computer programs. The aim of virtual screening is to identify molecules of novel chemical structure that bind to the macromolecular target of interest. Virtual screening is very important part of drug design and drug discovery research. [5][6][7][8] Drug designing is the inventive process of finding new medications based on the knowledge of a biological target. The drug is most commonly an organic small molecule that activates or inhibits the function of a bio-molecule such as a protein, which in turn results in a therapeutic benefit to the patient. Ligand-based drug design (or indirect drug design) relies on knowledge of other molecules that bind to the biological target of interest. QSAR relationships in turn may be used to predict the activity of new analogs. Structure-based drug design (or direct drug design) relies on knowledge of the three dimensional structure of the biological target obtained through methods such as x-ray crystallography or NMR spectroscopy. Some time we can do homology model of the target based on the experimental structure of a related protein. Computer-aided drug design uses computational chemistry to discover, enhance, or study drugs and related biologically active molecules. hit identification using virtual screening (structure or ligand based design), Hit-to-lead optimization of affinity and selectivity (structure-based design, QSAR, etc.) lead optimization of other pharmaceutical properties while maintaining affinity. [9][10][11][12][13]

MATERIALS AND METHODS

NCBI tools

Bioedit software- Version 7.1.3.0, Bioedit sequence alignment editor copyright (c) 1997-2011, Tom Hall. It is a sequence alignment programme, with this we can create a plasmid, restriction mapping, to know nucleotide composition and amino acid composition.

ORF – Open reading frame is any sequence of DNA or RNA that can be translated into a protein. In a gene, ORFs are located between the start code sequence (initiation codon) and the stop code sequence (termination codon). ORF tool is a graphical analysis tool which finds all open reading frames of selectable minimum size in a user sequence or in sequence which is already in database.

Map Viewer – The map viewer supports search and display genomic information by chromosomal position.

E-PCR – Electronic polymerase reaction is a computational procedure that is used to identify sequence tagged sites, with in DNA sequences.

VECSCREEN – VEC SCREEN is a system for quickly identifying segments of nucleic acid sequence that may be vector origin. NCBI developed VECSCREEN to combat the problem of vector contamination in public sequence databases.

Clustal w – It is used multiple sequence alignment computer program.

GENSCAN – Generally used to predict complete gene structures in human DNA and genomic DNA.

BLAST - Basic local alignment searching tool. BLAST uses a pair wise local search and uses a number of methods to increase the speed of the original Smith-waterman algorithm. Smith waterman Algorithm is a well known algorithm for performing local sequence alignment. That is for determining similar regions between two nucleotide or protein sequences.

MSA – Multiple sequence alignment: with this we can do a pair wise alignment, create phylogenetic tree (or use user define tree). Use the phylogenetic tree to carry out the multiple alignment.

Clustal W – Command line interface.

Phylodraw software

Phylodraw version 0.8.2, Graphic application lab, Pusan National University. By this software we can do Phylodraw Analysis. It represented in the form of trees. A tree is 2D graph showing evolutionary relationship. Types of trees like (rooted, unrooted, true, inferred, species, gene, phylogram, cladogram tree.)

Proteomics tools

Protparam - Physico-chemical parameters of a protein sequence (amino-acid and atomic compositions, isoelectric point, extinction coefficient, etc.)

Protscale - Amino acid scale representation (hydrophobicity, other conformational parameters, etc.)

GOR – The GOR method (Garnier Osguthrope Ribson) is an information theory based method for prediction of secondary structures in protein.

SOPMA –Sopma is secondary structure prediction method, self optimized prediction method with alignment.

SignalP - Prediction of signal peptide cleavage sites.

NetNGlyc - Prediction of N-glycosylation sites in human proteins.

NetOGlyc - Prediction of O-GalNAc (mucin type) glycosylation sites in mammalian proteins.

NetAcet - Prediction of N-acetyltransferase A (NatA) substrates (in yeast and mammalian proteins).

NetPhos - Prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Sulfinator Prediction of tyrosine sulfation sites

SOSUI – Prediction of transmembrane regions.

Active site analysis – The active site of an enzyme is the binding site where catalysis occurs. The structure and chemical properties of the active site allow the recognition and binding of the substrate.

Rasmol software - Version 2.7.5, it is a 3D analysis programme. The name RasMol comes from Raster display of molecules. Raster is a type of computer display especially useful for showing solid surfaces.

SPDBV software – Deep viewer/ Swiss pdb viewer v 3.7, Gsk Company, (c) 1995-2001 N-guex. SWISS PDV viewer is an application that provides a user friendly interface allowing analyzing several proteins at the same time.

VegaZZ Software – Version 2.4.0. It is also a molecular modelling toolkit. By the help of this software we can create a database and the database docking process is called virtual screening.

Argus Lab software – Argus Lab 4.0, copy right (c) 1997-2004, Mark Thompson and Planaria Software LLC, all right reserved, www.arguslab.com. It is a molecular modelling program.

DOCKING – It is very important part of drug discovery and drug designing. Docking is method which predicts and preferred orientation of one molecule to a second when bound to each other to form stable complex. [14][15]

Protein ligand docking- Protein – Ligand docking is a molecular modelling technique. The goal of protein ligand docking is to predict the position and orientation of a ligand, when it bound to a protein receptor or enzyme. [16]

Protein- protein docking – This involves two proteins that are approximately the same size. Therefore, usually the docking site is a more “planner” surface than in that

ligand-protein docking and cases where the docking occurs when one molecule is located inside a cavity in the other molecule are very rare.

HOMO - HIGHEST OCCUPIED MOLECULAR ORBITAL

It is the molecular orbital in a molecule that has the highest energy of all the occupied orbitals. An occupied orbital contains at least one electron.

LUMO – LOWEST UNOCUUPIED MOLECULAR ORBITAL

The lowest energy molecular orbital of an atom or molecule that does not contain an electron. If the atom or molecule were to accept an electron, it would be most likely to do it with this orbital .The HOMO is the orbital that could act as an electron donor, since it is the outermost (highest energy) orbital containing electrons. The LUMO is the orbital that could act as electron acceptor, since it is the innermost (lowest energy) orbital that has a room to accept electrons.

ESP MAPPED DENSITY – ELECTRO STATIC POTENTIAL MAPPED DENSITY

It is valuable for computer aided drug design because they help in optimization of electrostatic interactions between the protein and the ligand. These surfaces can be used to compare different inhibitors with substrates or transition states of the reaction. The molecular electrostatic potential is the potential energy of a proton at a particular location near molecule.

Hyperchem software – version 7.5.0

QSAR – QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP

QSAR properties (Gasteiger charges, surface area, volume, hydration energy, log p refractivity, sum of bond polarizabilities, mass).Geometry optimization a calculation, using either a molecular mechanics or semi-empirical method to find a minimum energy (stable) configuration for a molecular system. The calculation adjusts atomic coordinates in steps to find a configuration in which net forces on each atom are reduced to zero. QSAR represent an attempt to correlate structural or property descriptors of compounds with activities.

CAche software – COMPUTER AIDED CHEMISTRY

CAche enables to draw model molecules and perform calculations on a molecule to discover molecular properties and energy values. CAche displays experimental results in a variety of ways such as: Moving a molecule atoms and bonds to produce an optimized or low energy structure showing electronic properties as surfaces super imposed on a molecule producing 3D energy graphs viewed alongside a series of low-energy conformations.

CAche UV visible transitions

This indicates graphical representation, which comprises of X-axis as the wave length and Y-axis as the molar absorptivity.

CAche IR visible transitions

IR spectrum is a plot of wave number/ wave length (x-axis) vs. percent transmittance/absorbance (y-axis). By the help of this software without going to wet lab we can observe the maximum molar absorptivity (UV) & transmittance (IR) in our CAche graphical preview.

METHODS

Genomics

Genomic information from NCBI gene database of SNCA gene

Genbank format (to know the gene information)

Downloaded a FASTA format of SNCA gene (for our research purpose)

Bio-Edit

a. To know the nucleotide composition of SNCA gene.
b. Created a plasmid vector
c. And have done a restriction mapping

↓

NCBI tools

a. ORF (Open Reading Frame)
b. Map Viewer
c. e-PCR (electrostatic polymerase chain reaction)
d. Vec-Screen

↓

Gene Predicting Tools

a. Genscan

↓

Phylogenetic analysis

a. BLAST
b. FASTA
c. MSA (Multiple Sequence Alignment)
d. Clustal W
e. Phylodraw (for know the phylogenetic trees)

↓

Proteomics
Genpept format (to know protein information)
FASTA format of protein (for our research purpose)
Bioedit

a. Amino acid composition
b. Text
c. Graph

↓

Primary structure analysis

a. Protparam
b. Protscale

↓

Secondary structure analysis

↓

a. GOR
b. SOPMA

Post translational modifications

a. SignalP
b. NetNGly
c. NetOGly
d. NetAcet
e. NetPhos
f. Sulfinator

↓

Topology prediction

a. SOSUI

↓

SPDBV software

↓

a. Loading of raw sequence
b. After loading templates colouring of raw sequence and templates
c. Aligning of 1st template
d. Modelled protein before loop building
e. Ramchandran plot before loop building
f. Loop building (configuration table)
g. Modelled protein after loop building
h. Ramchandran plot after loop building
i. Protein with H bonds
j. Protein with side chain

↓

Active site analysis

a. Cavity method
b. Q-site finder

↓

3D analysis

a. RASMOL(software)

↓

DRUG

↓

a. Individual docking
b. Similar molecules
c. Database docking

↓

Argus Lab (surfaces) software

a. Open final molecule
b. HOMO (Highest occupied molecular orbital)
c. LUMO (lowest occupied molecular orbital)
d. ESP (Electro static potential) mapped density

↓

Hyperchem Software (QSAR properties)

a. Final molecule
b. Single point
c. Geometry optimization
d. Molecule with QSAR properties table
e. QSAR properties values

↓

CAche (computer aided chemistry) software

a. UV visible transitions and **b.** IR transitions

RESULT

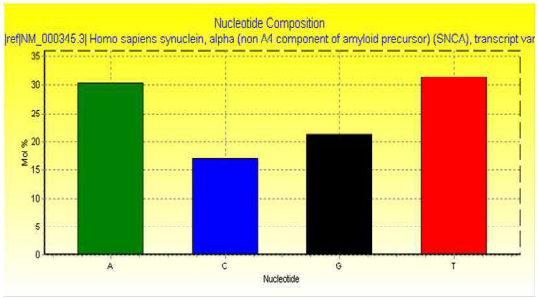


Fig 1: The Bio-Edit results of nucleotide composition shows that A+ T content is higher than G+C content with a score of 61.65%. It indicates that mol % of thymine i.e. 31. 32% and adenine having 30.33% was higher than the other nucleotide bases.

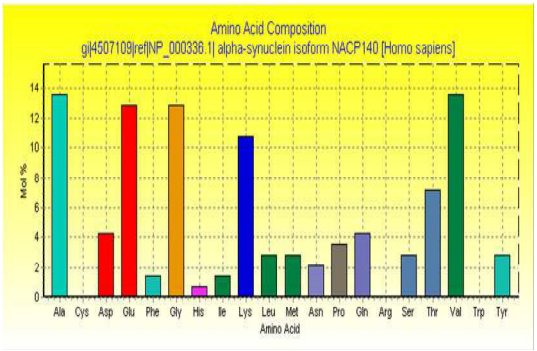


Fig 2: The bio-edit results of amino acid composition and it indicates that the mol % of alanine and valine were greater than all the other amino acid residues.

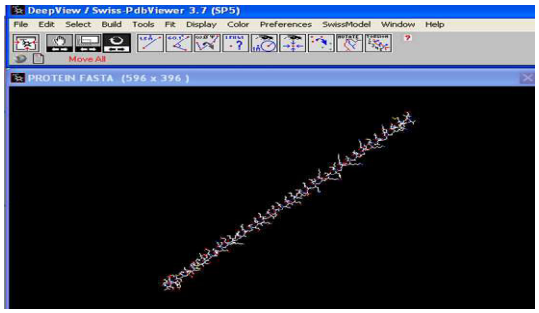


Fig 3: Protein FASTA format of alpha-synuclein isoform NACP140 [Homo sapiens].

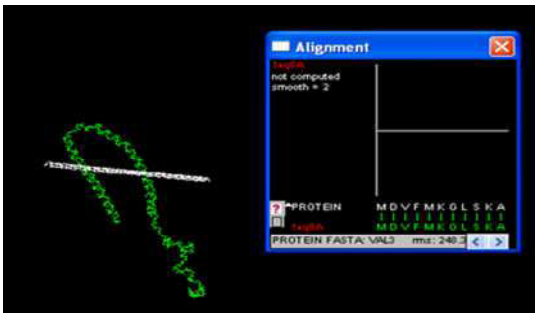


Fig 4: Alignment of FASTA format with 1xq8A template which is downloaded from swiss model.

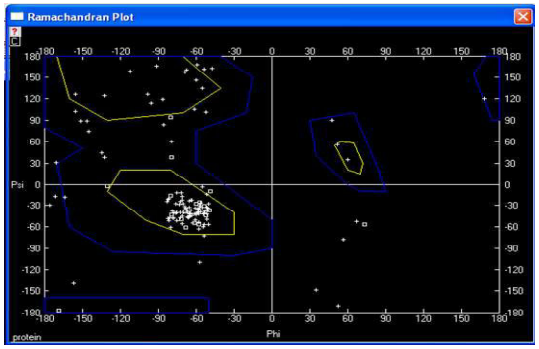


Fig 5: Ramchandran plot obtained through SPDBV software to know protein prediction and amino acids presence.

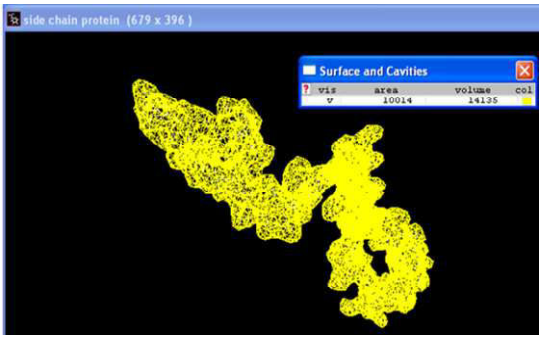


Fig 6: Protein with entire surface and cavities done through active site analysis.

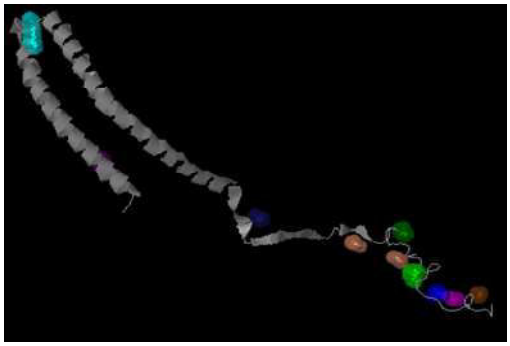


Fig 7: Active site analysis through Q-site finder method by Java application.

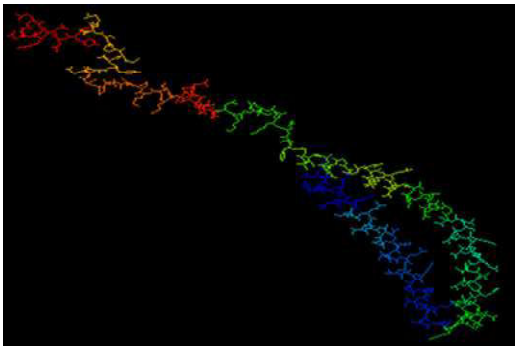


Fig 8: Protein side chain that is collected after completion of homology modelling.

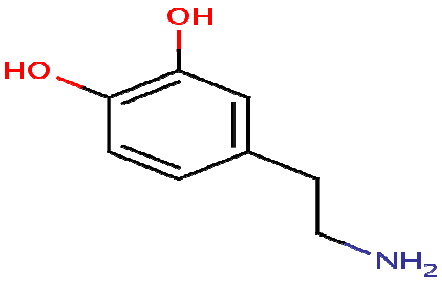


Fig 9: Dopamine Agonist structure that is downloaded from drug bank database.

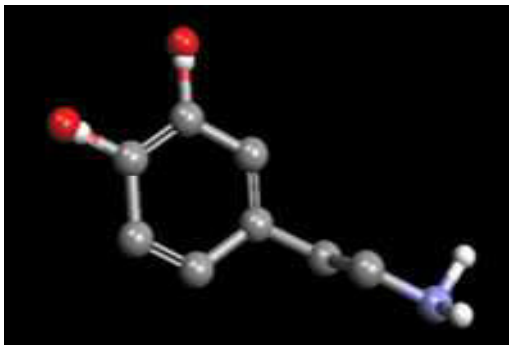


Fig 10: Dopamine Agonist which is designed and modelled through Argus lab software.

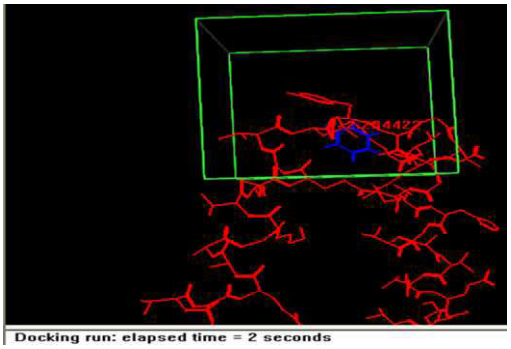


Fig 11: The Docking process for Dopamine Agonist (market available drug).

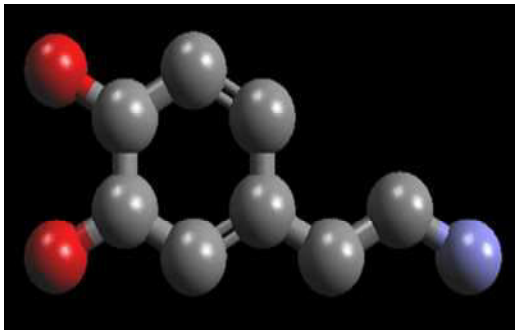


Fig 12: Similar drug molecules which are obtained from PUBCHEM database. This is the structure of Dopamine Quinone

Drugs name	Amino acid residue	Energy
Intropin	Tyr 125	-6.27895
Intropin	Glu 126	-6.34895
Dopamine quinone	Met 127	-6.34292
Dopamine quinone	Pro 128	-6.23941
Dopamine quinone	Ala 124	-6.20978
Dopamine quinone	Lys 96	-6.05365
Intropin	Val 118	-6.32469

Fig 13: Database docking done through Argus lab. The final drug was selected based on the least energy level and docking with more number of amino acids.

MET 127

Summary of results in order of docking score (kcal/mol)

****

1. Dopamine quinone -6.34292

2. Intropin -6.34292

3. 5, 6-Dihydroxyindole -5.88592

4. Selegiline -5.09018

5. 3, 4-Dihydroxyphenylacetic acid -4.63217

6. Etilevodopa -4.35275

7. Levodopa methyl ester -4.096

8. Sesquihydrate -2.05784

Elapsed time for calculation = 38 seconds

Fig 14: It has lowest energy level (MET 127). This is the result of database docking method.

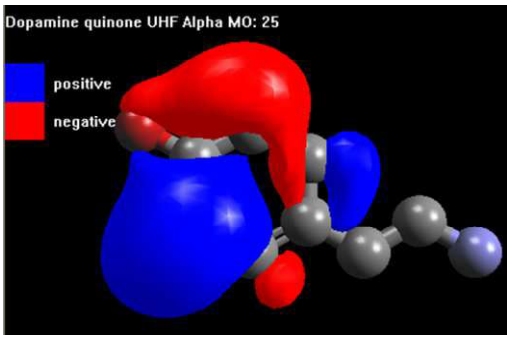


Fig 15: Quick plit HOMO (higher occupied molecular orbital).

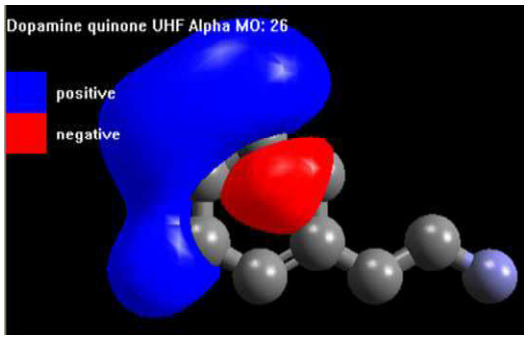


Fig 16: Quick plit LUMO (lowest unoccupied molecular orbital).

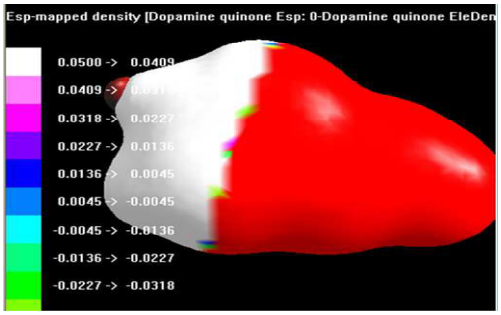
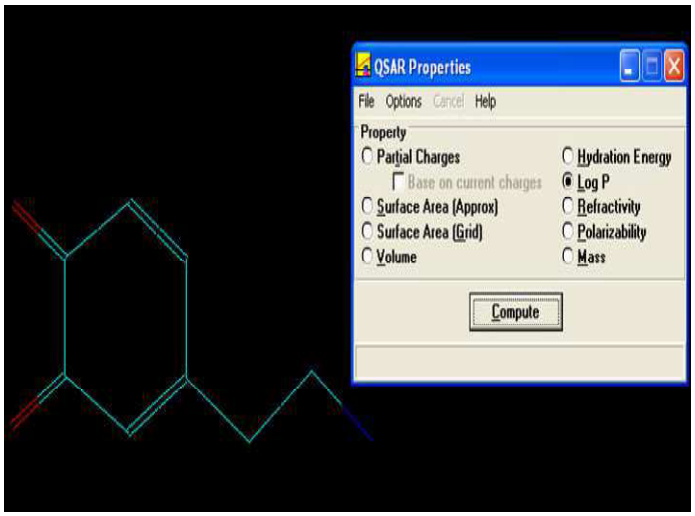


Fig 17: Quick plit ESP (electro static potential)



PARTIAL CHARGES: 0.00e

SURFACE AREA (APPROX) : 416.42A^2

SURFACE AREA (GRID): 338.27A^2

VOLUME: 505.30A^3

HYDROGEN ENERGY: -6.31kcal/mol

Log P: 1.47

REFRACTIVITY: 38.41A^3

POLARIZABILITY: 12.57A^3

MASS: 154.10 amu

Fig 18: Quantitative structure activity relationship (QSAR) is analysed in hyperchem software.

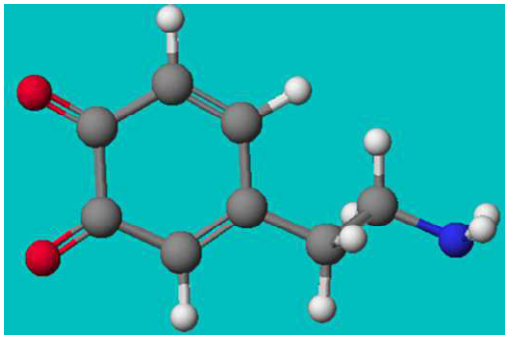


Fig 19: UV visible Dopamine Quinone structure in CAche.

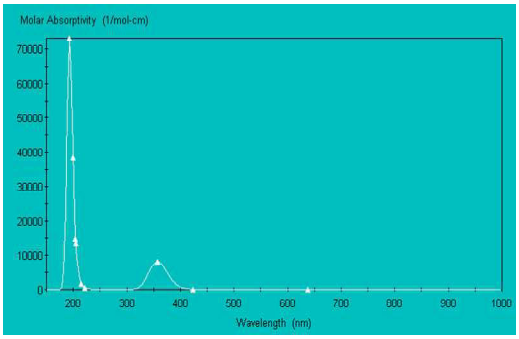


Fig 20: UV visible graph in CAche software.

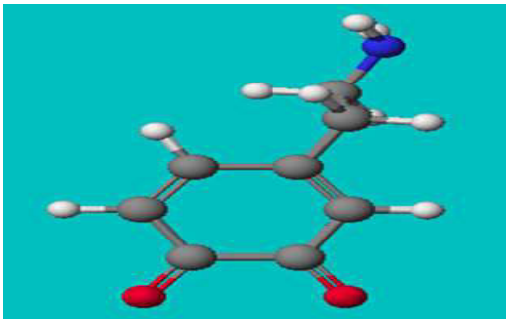


Fig 21: Infrared visible of Dopamine quinone in CAche software.

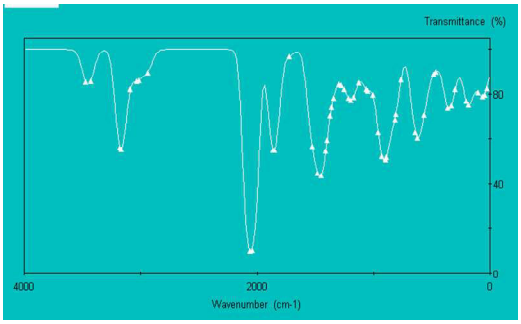


Fig 22: Infrared visible Graph in CAche software.

DISCUSSION

The protein and gene sequence of SNCA gene was retrieved from FASTA for Blast in NCBI server that served as a skeletal backbone for the identification of the SNCA gene in Parkinson's disease. The Bio-Edit results of nucleotide composition shows that A+T content is higher than G+C content with a score of 61.65% and indicates that mol % of thymine i.e. 31. 32% and adenine having 30.33% was higher than the other nucleotide bases which shown in Fig.1. The bio-edit results of amino acid composition indicate that the mol % of Alanine and Valine are greater than all the other amino acid residues shown in Fig.2. The primary structure analysis of alpha-synuclein protein shows in the Protparam, the aliphatic index is 69.64. Grand average of Hydropathicity (GRAVY): -0.403 shows hydrophilicity. Alpha synuclein is analysed through protscale to know molecular weight, bulkiness, polarity, recognition factor, number of codons, refractivity, HPLC of protein. Protein secondary structure analysis is done through GOR, SOPMA. The comparative homology modelling of protein structure is done through SPDBV (Swiss pdb viewer) software. Protein FASTA format of alpha-synuclein isoform NACP140 Homo sapiens is opened in SPDBV software, shown in Fig.3. Alignment of protein FASTA format with 1xq8A template, which is downloaded from Swiss model proteomics tool shown in Fig.4. Ramchandran plot obtained through SPDBV software to know protein prediction and amino acids shown in Fig.5. The energy minimization was done for optimization with the help of side chain proteins. The highest volume surface of the protein was computed in the cavity method, which shown in Fig.6. The active site analysis is done through Q-site finder method and 3D structure shown in JAVA application, Shown in Fig.7. Protein side chain is collected after completion of homology modelling in SPDBV software in Fig.8. Standard market available drugs such as Carbidopa, Dopamine, Levodopa, Memantine, and Apomorphine were retrieved from drug bank database, one example (Dopamine Agonist) is shown in Fig.9. Drug molecules are designed and modelled in Argus lab software, one example (Dopamine Agonist) is shown in Fig.10. The individual docking of the drugs were done with amino acids from Q-site finder. The docking process of Dopamine Agonist is done, shown in Fig.11. Similar drug molecules for each drug molecules were retrieved from PUBCHEM database of NCBI server, which are approximately similar to molecular weight and chemical identification number (CID) such as 1,2,3,4-Tetrahydroisoquinoline; 2-hydrazinyl-2-methylpropanoate; 3-(2-methylpropyl) adamantan-1-amine;3,4 ihydroxyphenylacetic acid; 3-Benzoyl dopamine; 3-Benzoyl dopamine; 3-propyladamantan-1-amine;5,6 Dihydroxyindole; Alpha methyldopa; Dopamine quinone; Etilevodopa; Intropin; Levodopa methyl ester; methyl 3-hydroxy-alpha-methyl-L-tyrosinate; Selegiline; Sesquihydrate. Database is created by Vega ZZ software retrieving above similar drug molecules. Each drug molecules are designed in Argus lab software one example (Dopamine Quinone) is shown in Fig.12. Database docking done through Argus lab. The final drug was selected based on the least energy level and docking with more number of amino acids, shown in Fig.13. According to database docking result Met 127 has lowest energy level. Shown in Fig.14. It was concluded that Dopamine Quinone is docked 17 times with this amino acid residues i.e. Lys 32, Glu 35, Leu 38, Val 40, Gly 41, Thr 44, Asp 121, Ala 124, Met 127, Pro 128, Glu114, Met 116, Pro 117, Ser 129, Glu 130, Gly 132 and Lys 96. The final molecule Dopamine Quinone was confirmed as it has the lowest energy and the highest time repetition in database docking. HOMO (higher occupied molecular orbital), LUMO (lowest unoccupied molecular orbital), ESP (electro static potential) of Dopamine Quinone are done through Argus lab software, shown in Fig.15.16.17. UV visible, UV visible graph, infrared visible, infrared visible graph of Dopamine Quione in CAche software, shown in Fig.19.20.21.22. The final molecule are computed in hyperchem QSAR, which shows parameters like partial charges = 0.00e, surface area (approx) = 416.42A^2, surface area (grid) = 338.27A^2, volume = 505.30A^3,

hydrogen energy = -6.31kcal/mol, Log P =1.47, refractivity = 38.41A^3, polarizability = 12.57A^3, mass = 154.10 amu, which are shown in Fig.18. The final molecule Dopmaine Quinone was identified with lowest energy minimization i.e. -6.34292 and it was cleared that this protein molecule became more stable after energy minimization.

CONCLUSION

The results obtained that Dopamine Quinone is interacting at the lowest energy level with all the amino acids in the potential active site. This Dopamine Quinone is highly similar to our market available drug that is Dopamine Agonist. Thus Dopamine Quinone drug molecule can be a medicine for Parkinson's disease in the near future. Analyzing of protein and gene information can stop the mutation in SNCA gene by the process of protein engineering with the help of bioinformatics and biotechnology.

REFERENCES

1. Jankovic J, Parkinson's disease: clinical features and diagnosis. Journal of neurology, neuro surgery, psychiatry 79: 368-376 (2008).
2. Barnett-Cowan M, Dyde RT, Fox SH, Moro E, Hutchison WD, Harris LR, multisensory determinants of orientation perception in Parkinson's disease: Neuroscience,167(4): 1138-50, (Jun 2010).
3. Nuytemans K, Theuns J, Cruts M, Van Broeckhoven C., Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes; a mutation update: Human mutation, 31(7):763-80, (2010).
4. Wassilios G Meissner, Mark Frasier, Thomas Gasser, Christopher G Goetz, Andres Lozano, Paola Piccini, José A Obeso, Olivier Rascol, Anthony Schapira, Valerie Voon, David M Weiner, François Tison & Erwan Bezard. Priorities in Parkinson's disease research: Nature Reviews Drug Discovery, 10: 377-393, (2011).
5. Rester U., from virtuality to reality - Virtual screening in lead discovery and lead optimization: a medicinal chemistry perspective. Curr Opin Drug Discov Devel, 11(4): 559-68, (July 2008).
6. Patrick Walters W., Matthew T Stahl, Mark A Murcko, Virtual screening-an overview: Drug Discovery Today, 3: 160-178, (April 1998).
7. Srinivas Reddy A, Priyadarshini Pati S, Praveen Kumar P, Pradeep HN and Narahari G Sastry. Virtual Screening in Drug Discovery – A Computational Perspective Current Protein and Peptide Science, 8: 329-351 (2007).
8. Tingjun Hou and Xiaojie Xu., Recent Development and Application of Virtual Screening in Drug Discovery: Current Pharmaceutical Design, 10: 1011-1033, (2004).
9. Tollenaere JP., the role of structure-based ligand design and molecular modelling in drug discovery: Pharmacy and world sciences, 18(2): 56-62, (Apr 1996).
10. Renxiao Wang, Ying Gao and Luhua Lai, LigBuilder: A Multi-Purpose Program for Structure-Based Drug Design, Journal of Molecular Modelling, 6: 498-516, (2000).
11. Jorgensen WL., the many roles of computation in drug discovery, Science, 303: (5665), Mar 19, 2004.
12. F. Ooms*, Molecular Modelling and Computer Aided Drug Design. Examples of their Applications in Medicinal Chemistry: Current Medicinal Chemistry, 7:141-158, (2000).
13. Facultés F. Universitaire, Notre-Dame de la Paix. Molecular Modeling and Computer Aided Drug Design. Examples of their Applications in Medicinal

Chemistry: Current Medicinal Chemistry, 7: 141-158, (2000).
14. Kitchen DB, Decornez H, Furr JR, Bajorath J., Docking and scoring in virtual screening for drug discovery: methods and applications, Nature Review drug discovery, 3(11): 935-49, (Nov 2004).
15. Lengauer T, Rarey M, Computational methods for bio molecular docking: Current opinion in structural biology, 6(3): 402-406, (Jun 1996).
16. Cerqueira NM, Bras NF, Fernandes PA, Ramos MJ, MADAMM: a multistaged docking with an automated molecular modelling protocol: Proteins, 74(1): 192-206 (Jan 2009).