



Docking studies of Alzheimer Cathepsin B protein-ligand complex in presence and absence of water molecules: Insight to enhance docking accuracy of AutoDock4.0

*¹Hashnu Dutta

*¹Amity Institute of Biotechnology, Amity University, Noida, U.P
Phone:9910963292, Email: teeyash24@gmail.com

Received date: Nov 6th 2012, Accepted date: Nov 15th 2012, Published date: Jan 1st 2013

Abstract:

Alzheimer's disease (AD) or Senile Dementia of the Alzheimer Type (SDAT) is a progressive neurologic brain disease leading to irreversible loss of neurons and loss of intellectual abilities, including memory and reasoning, which impedes social or occupational functioning. During the course of the disease, *plaques* and *tangles* develop within the brain. This causes neuronal brain death. AD patients also have a deficiency in the levels of neurotransmitters which are involved with the transmission of messages in the brain. **Cathepsin B** in secretory vesicles has been recently identified as a β -secretase for the production of neurotoxic β -amyloid ($A\beta$) peptides that accumulate causing severe memory loss in AD. Protein Cathepsin B with its certain ligands was selected that bind to the protein and inhibit its action. The protein and the ligands were optimized, docked and their interaction was visualized using UCSF Chimera. The result obtained was analyzed based on various factors such as hydrogen bonds, estimated binding free energy and K_i value. The results demonstrate that the accuracy of determination of complex geometry using AutoDock 4 for docking calculation greatly increases with the use of water molecules in the binding pocket of protein.

Keywords: Docking, Cathepsin B, Auto-Dock 4, Alzheimer's, Chimera, Plaques.

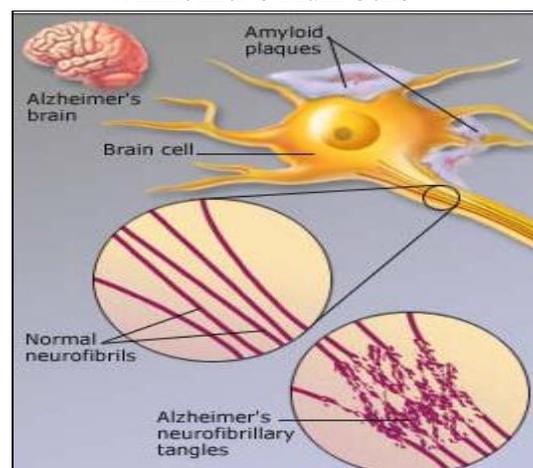
Introduction:

Background

When proteins are in solution, their entire surfaces are covered by water molecules with properties different from bulk properties. Most of the water molecules in the vicinity of a protein are loosely bound to it. The presence of water molecules plays an important role in the accuracy of ligand-protein docking predictions. The prediction of binding modes and energies of ligands to proteins by docking are commonly employed from several years in drug designing. In general, the water molecules are removed prior to dock a ligand in Auto-Dock software. To find the accuracy of the complex geometry and binding energy, the water molecules are included and removed in the same binding pocket of set of Alzheimer Cathepsin B (CatB) protein ligand complex.

247

Alzheimer's Brain Cells



Materials and Methods:

Various Important Databases

- **PDB (Protein Data Bank):** Repository of atomic co-ordinates and three-dimensional structural data of important biological macromolecules, such as, proteins and nucleic acids, maintained in Rutgers and UCSD.
- **PubChem:** Database of chemical molecules and their activities against biological assays, maintained by Nation Center for Biotechnology Information (NCBI) in Bethesda, Maryland.
- **KEGG (Kyoto Encyclopaedia of Genes and Genomes):** Bioinformatics resource for linking genomes to life and the environment. It is a collection of online databases dealing with genomes, enzymatic pathways, and biological chemicals, which is maintained in Japan.

Various Softwares Used

1. **ChemSketch:** allows drawing of chemical structures including organics, organometallics, polymers, and Markush structures. It also includes features such as calculation of molecular properties and logP prediction.
2. **UCSF Chimera:** highly extensible program for interactive visualization and analysis of molecular structures.
3. **AutoDock 1.5.4:** helps to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure.

4. Cygwin Terminal: provides native integration of Windows-based applications, data, and other system resources with applications, software tools, and data of the Unix-like environment.

5. Open Babel GUI: converts chemical objects (currently molecules or reactions) from one file format to another.

Methodology:

Different compounds of interest were selected and retrieved from PDB data bank. Protein.pdb files for each of the compounds were prepared along with their respective ligands. The pdb structures are subjected to optimization both in the presence and absence of water molecules using UCSF Chimera software. The structures of the ligands used for the treatment of Alzheimer's disease were prepared using Draw mode of ACD/ChemSketch software. The pdb files were then converted to mol format using open babel software. The physio-chemical properties of the respective ligands were calculated using Structure mode of ChemSketch. LogP values were calculated using ChemSketch. Using AutoDock software and Autogrid option, protein.pdbqt, ligand.pdbqt, protein.gpf, and protein.dpf format files were prepared; polar hydrogen and gasteiger charges were assigned to ligands by different optimization techniques. Docking procedure used was Cygwin software using the two important commands:

Autogrid command: `./autogrid4.exe -p protein.gpf -l protein.glg &`

Autodock command: `./autodock4.exe -p protein.dpf -l protein.dlg &`

Finally *.glg and *.dlg format files were created. Files with protein_docked.pdb formats were created for identifying lowest RMSD, and different types of energy values.

Results and Discussion:

RMSD Table summarizes the efficacy of the docking search method in the form of the average best RMSD obtained for each protein, clearly revealing that in both the absence and the presence of water molecules, the docking search method was highly successful in finding binding poses close to the crystallographic binding modes (the highest mean RMSD is **3.215 Å** for docking simulations of the Cathepsin B binding protein in the presence of water molecules), with only a marginal improvement of the overall average best RMSD from **3.215-3.369 Å**. The overall average RMSD of the top-ranked binding poses improved in the presence of water molecules from **3.4 Å to 3.2 Å**, with the difference shown to be statistically significant with a paired *t*-test (*p*-value= 0.001355). These results show that the accuracy improved significantly in the presence of water molecules for all 17 Cathepsin B, with the

improvement being more noticeable for PDB ID **3AI8**.

It also summarizes all cross-docking simulations by reporting the number of ligands that were successfully docked with each protein. Successful docking simulations were considered to be those whose top-ranked binding pose was within 2.2 Å RMSD of the crystallographic binding mode. A success rate of **35% (17 ligands out of 47)** was achieved in the absence of water molecules (No W), increasing substantially to **72% (34 ligands out of 47)** in the presence of water molecules (W).

Solvated docking leads to improvements both in quality and scoring. The presence of water molecules plays an important role in the accuracy of ligand-protein docking predictions. We have observed a statistically significant overall increase in accuracy when water molecules are included during docking simulations and have found this to be independent of the method of optimization of the orientation of water molecules. These results confirm the importance of including water molecules whenever possible in a ligand-protein docking simulation. Findings also reveal that prior optimization of the orientation of water molecules, in the absence of any bound ligand, does not have a detrimental effect on the improved accuracy of ligand-protein docking. This is important, given the use of docking simulations to predict the binding modes of new ligands or drug molecules.

Physiochemical Properties Of Screened Compounds:

Compound Name	*Log P
HNQ	2
E6C	2.03
O74	0.54
EP2	-0.03
73V	2.2
O42	0.47
59A	0.4
74M	1.4
75V	0.2
76V	2.11
77B	2.98
78A	1
PYS	-0.16
EPO	0.32
NAG	-2.45
OE6	3.01

* logP: Partition coefficient

Protein PDB Ids With Their Chosen Ligands

PDB IDs	Ligands With Their Codes
3AI8	5-nitroquinolin-8-ol [HNQ]
1ITO	N-[1-hydroxycarboxyethyl-carbonyl]leucylamino-2-methyl-butane [E6C]
1QDQ	[propylamino-3-hydroxy-butan-1,4-dionyl]- isoleucyl-proline [O74]
1SP4	methyl N-[(2S)-4-{{(1S)-1-{{(2S)-2-carboxypyrrolidin-1-yl}carbonyl}-3-methylbutyl}amino)-2-hydroxy-4-oxobutanoyl]-L-leucylglycylglycinate [EP2]
2DC6	benzyl N-((2s,3s)-3-[(propylamino)carbonyl]oxiran-2-yl)carbonyl)-L-isoleucyl-L-prolinate [73V]
2DC7	N-[[{(2s,3s)-3-(ethoxycarbonyl)oxiran-2-yl}carbonyl]-L-threonyl-L-isoleucine [O42]
2DC8	N-[[{(2s,3s)-3-(ethoxycarbonyl)oxiran-2-yl}carbonyl]-L-isoleucine [59A]
2DC9	methyl N-((2s,3s)-3-[(propylamino)carbonyl]oxiran-2-yl)carbonyl)-L-isoleucyl-L-prolinate [74M]
2DCA	N-[[{(2s,3s)-3-(ethoxycarbonyl)oxiran-2-yl}carbonyl]-L-isoleucyl-L-alanine [75V]
2DCB	N-[[{(2s,3s)-3-(ethoxycarbonyl)oxiran-2-yl}carbonyl]-L-isoleucyl-L-isoleucine [76V]
2DCC	benzyl N-((2s,3s)-3-[(benzylamino)carbonyl]oxiran-2-yl)carbonyl)-L-isoleucyl-L-prolinate [77B]
2DCD	N-((2s,3s)-3-[(benzylamino)carbonyl]oxiran-2-yl)carbonyl)-L-isoleucyl-L-proline [78A]
1CTE	2-pyridinethiol [PYS]
1CSB	N-[(3R)-4-ethoxy-3-hydroxy-4-oxobutanoyl]-L-isoleucyl-L-proline [EPO]
2IPP	2-pyridinethiol [PYS]
3MOR	N-acetyl D-glucosamine [NAG]
1THE	amino{[(4S)-4-{{(benzyloxy)carbonyl}amino}-5-{{(2S)-1-(benzyloxy)-4-chloro-3-oxobutan-2-yl}amino)-5-oxopentyl}amino}methaniminium [OE6]

Calculated Energies With Rmsd Values Of Proteins In Presence of Water

PDB ID	E ^b	V ^{dw} -H ^b -D ^s	I ^c	T ^d	BE ^e	Ki (mM)	RMSD
3AI8	0.35	-5.34	-0.1	0.27	-4.81	0.3	1.383
1ITO	0.45	-6.57	-1.86	3.02	-4.96	0.23	1.695
1QDQ	0.5	-7.85	-0.98	2.74	-5.59	0.08	1.773
1SP4	0.77	-8.22	-1.7	4.39	-4.75	0.33	2.398
2DC6	-0.08	-8.53	-1.52	3.29	-6.84	0.01	4.517
2DC7	0.17	-7.64	-0.9	3.02	-5.35	0.01	3.083
2DC8	0.4	-6.34	-0.42	2.2	-4.17	0.88	1.473
2DC9	-0.06	-8.63	-1.21	2.74	-7.15	0.01	4.01
2DCA	0.45	-6.02	-0.96	2.74	-3.78	1.7	3.565
2DCB	0.37	-7.37	-1.23	3.29	-4.94	0.24	3.236
2DCC	0.02	-7.97	-2.15	3.29	-6.8	0.01	5.72
2DCD	0.41	-8.68	-1.23	2.47	-7.03	0.01	2.26
1CTE	0	-4.1	0	0	-4.1	0.99	0.813
1CSB	0.46	-7.72	-1.2	2.74	-5.72	0.07	1.672
2IPP	-0.06	-3.31	0	0	-3.38	3.35	10.019
3MOR	-0.05	-7.65	-1.39	1.65	-7.44	0.004	2.94
1THE	-1.18	-7.43	-2.26	4.12	-6.76	0.01	4.099

In Absence of Water

PDB ID	E ^b	V ^{dw} -H ^b -D ^s	I ^c	T ^d	BE ^e	Ki (mM)	RMS D
3AI8	0.38	-5.36	-0.1	0.27	-4.8	0.3	1.43
1ITO	0.25	-6.2	-0.72	2.74	-3.93	1.32	2.281
1QDQ	0.71	-7.44	-1.65	2.74	-5.64	0.07	2.074
1SP4	0.61	-8.15	-2.3	4.39	-5.45	0.1	3.281
2DC6	-0.17	-8.25	-1.48	3.29	-6.6	0.02	3.74
2DC7	0.46	-6.77	-0.64	3.02	-3.93	1.33	3.156
2DC8	0.19	-6.77	-0.41	2.2	-4.81	0.3	3.151
2DC9	0.02	-7.85	-1.15	2.74	-6.24	0.03	4.102
2DCA	0.5	-5.96	-1.05	2.74	-3.76	1.76	3.631
2DCB	0.41	-7.37	-1.51	3.29	-5.18	0.16	2.991
2DCC	-0.06	-8.85	-1.49	3.29	-7.11	0.01	4.138
2DCD	0.38	-8.14	-1.3	2.47	-6.58	0.02	1.867
1CTE	0	-3.95	0	0	-3.94	1.29	1.337
1CSB	0.62	-6.96	-1.41	2.74	-5.01	0.21	4.314
2IPP	-0.07	-3.19	0	0	-3.25	4.12	10.214
3MOR	-0.08	-6.96	-1.45	1.65	-6.84	0.01	3.146
1THE	-0.63	-7.27	-1.83	4.12	-5.62	0.08	2.424

E^b = The electrostatic component of binding free energy in kcal/mol.

V^{dw}-H^b-D^s = Vander waals, Hydrogen Bond-Desolvation Energy Component of Binding Free Energy (kcal/mol)

I^c = Total Internal Energy (kcal/mol)

T^d = Torsional Energy of Binding (kcal/mol)

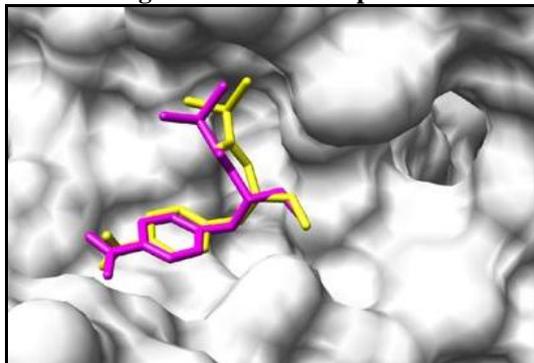
BE^e = Estimated Binding Free Energy (kcal/mol)

Ki (mM) = Inhibitory Constant (mM)

RMSD = Root Mean Square Deviation (Å)

Minimum RMSD value was chosen considering 'centre of the ligand' while Autodocking, Grid Size= 60*60*60, Spacing= 0.375 (Å).

Docked Ligand with the Receptor



Conclusion:

"Cathepsin B" protein PDB structure from RCSB was selected and its optimization was done by using **UCSF Chimera** and the active site of protein was found using Ligand Explorer. Ligand dataset of 17 compounds was selected from PDB. The structure of ligands were sketched using **ChemSketch** converted to *.pdb using **Open Babel** and optimized by UCSF Chimera. Molecular Docking of protein and ligands was performed using **AutoDock 4.2** in presence & absence of water molecules. Interactional analysis of protein-ligand complex was performed using UCSF Chimera again. The studies conclude that water is the best solvent involved in enhancing the binding affinity of the Ligand with the receptor since the docking accuracy was found to be more in the presence of water rather than in its absence.

Acknowledgement:

I would like to thank the Faculty-in-charge of this Innovative Project, who helped me learn the basics in details, with ease and perfection and in no time I was able to obtain optimum efficient results.

Also, my token of appreciation to the DNA Conference Technical Programme Committee 2012, which provided me a platform to present my project report in a formal manner putting some more informative tidbits and making it a presentable one. Last, but not the least, my Institute, Amity University, Noida, for helping me gain this valuable knowledge experience.

References:

1. Cheng A, Diller DJ, Dixon SL, Egan WJ, Lauri G, Merz KMJ., Jr Computation of the physico-chemical properties and data mining of large molecular collections. *J Comput Chem.*2002;23:172–183.
2. Aisen PS, Schneider LS, Sano M, Diaz-Arrastia R, van Dyck CH, et al. High-dose B vitamin supplementation and cognitive decline in Alzheimer's disease: a randomized controlled trial. *JAMA.* 2008;300:1774-1783.
3. Ramasamy Thilagavathi, Ricardo L. Mancera et al (2010). "Ligand-protein cross-docking with water molecules". *J. Chem. Inf. Model.*; 50:415-421.
4. Knopman DS. Alzheimer's disease and other dementias. In: Goldman L, Schafer AI, eds. *Cecil Medicine.* 24th ed. Philadelphia, Pa: Saunders Elsevier; 2011:chap 409.
5. Peterson RC. Clinical practice. Mild cognitive impairment. *N Engl J Med* 2011 Jun 9;364(23):2227-2234.
6. Mueller-Steiner S, Zhou Y, Arai H, Roberson ED, Sun B, Chen J, Wang X, Yu G, Esposito L, Mucke L, Gan L (September 2006). "Anti-amyloidogenic and neuroprotective functions of cathepsin B: implications for Alzheimer's disease". *Neuron* 51 (6): 703–14.
7. Pol E, Björk I (September 2001). "Role of the single cysteine residue, Cys 3, of human and bovine cystatin B (stefin B) in the inhibition of cysteine proteinases". *Protein Sci.* 10 (9): 1729–38.
8. Estrada S, Nycander M, Hill NJ, Craven CJ, Waltho JP, Björk I (May 1998). "The role of Gly-4 of human cystatin A (stefin A) in the binding of target proteinases. Characterization by kinetic and equilibrium methods of the interactions of cystatin A Gly-4 mutants with papain, cathepsin B, and cathepsin L". *Biochemistry* 37 (20): 7551–60.
