The Third Annual Conference of the Korean Society of Bioinformatics (KSBI2004)

Efficiency of Lamarckian Genetic Algorithm in Molecular Docking of Phenylaminopyrimidine (PAP) Derivatives: A Retrospect Study

Amor A. San Juan[1]*, Eva Marie A. Ratilla[2]

[1] Biochemical Research Center, Life Science Division, Korea Institute of Science and Technology, Seoul, Korea
[2] Institute of Chemistry, College of Science, University of the Philippines Diliman, Quezon City, Philippines

* To whom correspondence should be addressed. Email: asanjuan@kist.re.kr

Abstract

Molecular docking using Lamarckian genetic algorithm of AutoDock 3.0 (AD3) was employed to understand in retrospect the selectivity of phenylaminopyrimidine (PAP) derivatives against the kinase domain c-Abl, implicated in chronic myelogenous leukemia (CML). The energetics of protein-ligand complex was scored using AD3 to identify active drug conformations while Ligplot and ligand protein contact (LPC) programs were used to probe schematic molecular recognition of the bound inhibitor to the protein. Results signify correlation between model and crystal structures of STI-571 compound or Imatinib (IM), a PAP derivative and now clinically proven for its efficacy in CML. A prospect active form Abl inhibitor scaffold from matlystatin class of compounds will be published elsewhere.

Introduction

Computational docking is a rapidly growing tool in drug design. This technique yields drug model built on available protein crystal structure and conformational coordinates of inhibitor compounds. Through the use of AutoDock3 software, the computational binding energy of ligand-protein complex is employed to generate a lead compound. AutoDock program was developed at The Scripps Research Institute (La Jolla California) to provide an automated docking procedure in predicting ligand interaction to biomacromolecular targets.

A docking procedure consists of search algorithm and scoring function. Search algorithm is employed to effectively sample the 3D space of ligand and protein complex. Older version of AutoDock used simulated annealing where initial temperature of 39 ^{0}C was taken then gradually reduced after each cycle until global minimum is obtained. In contrast, current version of AutoDock used Lamarckian genetic algorithm (LGA) with both genetic algorithm (for global search) and Solis and Wets algorithm for local search.

Computational docking based on LGA [1] of AD3 was employed in this study. The flowchart of energy optimization by LGA is shown in Figure 1. First, copies of ligands were randomly created. Next, the fitness was evaluated by calculating the lowest docked

conformation of ligand-protein complex. The selection operator then chooses which ligands will reproduce. Subsequently, crossover and mutation were applied to generate new ligand positions and conformations. The resulting elite individual undergoes local search (LS) optimization. LS is done before each GA generation (except for the initial one). Then, the whole process starts back again to fitness evaluation until convergence or number of steps are met.

LGA PROCESS

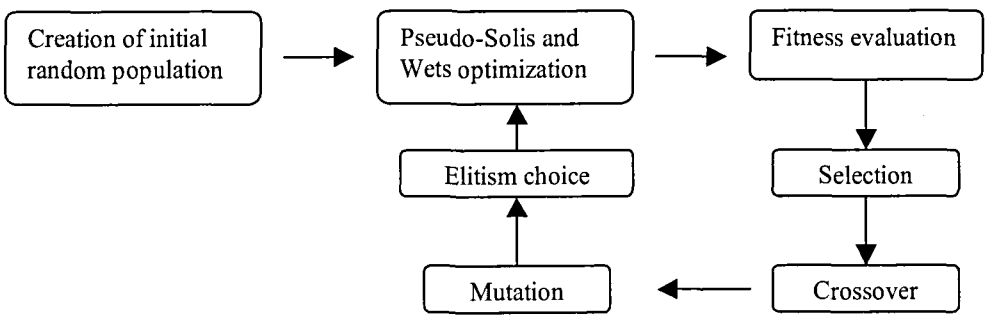


Figure 1. Flowchart of energy optimization process by LGA.

Chronic myelogenous leukemia (CML) is a fatal myeloproliferative disorder that progresses through three phases: chronic phase, accelerated phase and blast crisis. The molecular cause [2] of CML is the chimeric Bcr-Abl oncoprotein (contained in 95% of CML patients) that results from Philadelphia chromosome, a translocated chromosome from fused BCR-ABL genes.

Abl non-receptor tyrosine kinase is involved in cellular processes however, its structural alteration leads to human leukemias. The oncogenic activity of the Bcr-Abl depends on the Bcr fusion to Abl that results to active c-Abl transformation shown in cancer cells formation. The kinase domain of Abl has been crystallized into two conformations when bound to inhibitors. The active form Abl is characterized by open activation loop [3] in which Tyr 393 is phosphorylated while the inactive Abl form has closed activation loop that remains unphosphorylated. The inactive form of Abl consists of 293 residues in each chain A and B with 8 alpha and 9 beta helices. Similarly, the active form of Abl contains 293 residues in both chains A and B however, there are 9 alpha chains and 12 beta chains present.

A breakthrough drug Gleevec (also called Imatinib or STI-571) is an antiproliferative agent and the first signal transduction inhibitor for CML that effectively inhibits Bcr-Abl oncoprotein [4]. STI-571 is an Abl specific tyrosine kinase inhibitor that targets ATP binding site blocking protein phosphorylation that stops the pathways signaling tumor growth. Inhibition of Bcr-Abl kinase activity results [5] to transcriptional modulation of various genes involved in the control of cell cycle, cell adhesion and cytoskeleton organization.

Clinically tested therapeutic STI-571 drug had shown [6] resistance to CML blast-crisis patients. Findings revealed [6] that Abl-kinase oncoprotein shows dynamic behavior by its capability to switch into both active and inactive conformations. STI-571 was proven [7] by biochemical assay to have preferred binding for inactive Abl kinase loop than for active form. The design of potent inhibitor against active Abl form is currently being investigated.

Materials and Methods

Molecular modeling was performed on a Linux operating system (Red Hat version 8.0) equipped with 950MHz computer processor. Spartan Linux 2002 (Wavefunction Inc., Irvine, California) was used for molecular building and optimization. A complete development environment installation was employed to both AD3 and Ligplot source packages programs. The public domain AutoDock version 3.0 (Molecular Graphics Laboratory, The Scripps Research Institute, La Jolla California) was employed for computational docking of ligands to macromolecule. Ligplot version 4.0 (Department of Biochemistry and Molecular Biology, University College London) further analyzed the bound ligand-protein coordinates to show the plot of interatomic distances in between atoms within the active site. Ligand Protein Contact (LPC) software (Weizmann Institute of Science, Israel) analyzed the contact surface area between the ligand and protein. Lastly, SPSS 11 (SPSS Business Intelligence Division, Singapore) executed the data statiscal analysis.

The flowchart of computational methods performed in this study is shown in Figure 2 below.

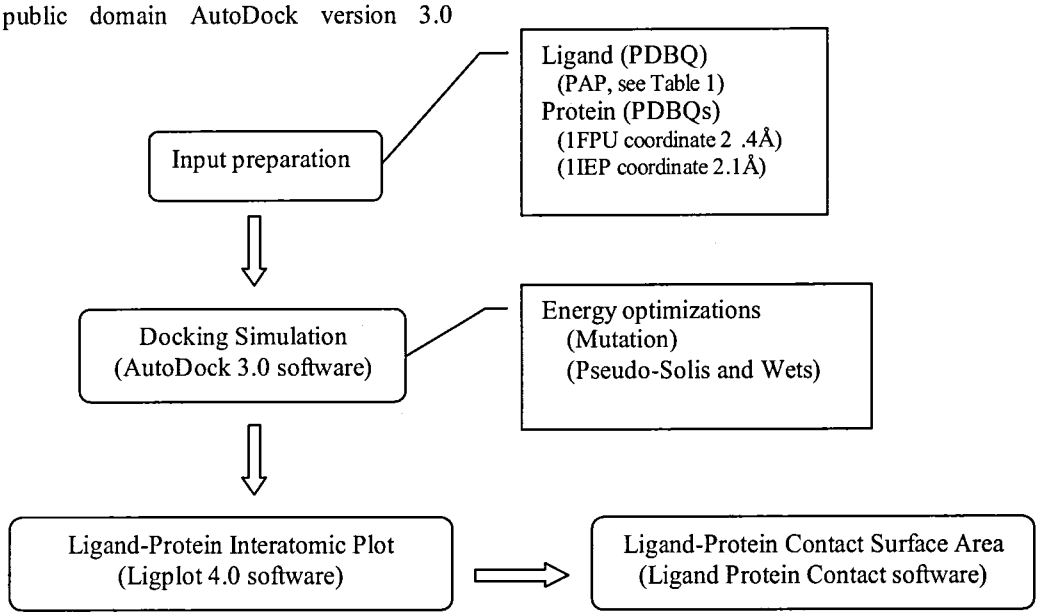
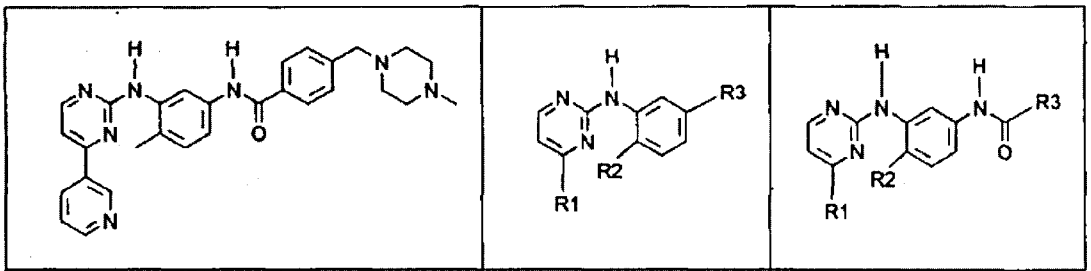


Figure 2. Biomodeling schematic diagram of CML structure-based drug design.

PDBQ = Protein data bank coordinate with atomic charges of ligand

PDBQs = Protein data bank coordinate with atomic charges and solvation of ligand



1	A	B

No.	Form.	R1	R2	R3
1	1	see above	"	"
2	B	3-pyridyl	H	identical as in compound 1
3	B	3-pyridyl	H	3-pyridyl
4	B	3-pyridyl	CH_3	4-methylphenyl
5	B	3-pyridyl	H	pentyl
6	B	3-pyridyl	CH_3	2-naphtyl
7	B	3-pyridyl	H	4-fluorophenyl
8	B	3-pyridyl	H	2-thiophenyl
9	A	see scheme 2	"	tetrafluoroethoxy
10	A	3-indolyl	H	tetrafluoroethoxy
11	B	3-pyridyl	H	phenyl
12	B	3-pyridyl	CH_3	phenyl
13	A	3-pyridyl	H	3-aminopropylaminocarbonyl
14	B	3-pyridyl	H	cyclohexyl
15	B	3-pyridyl	H	4-pyridyl
16	A	4-pyridyl	H	amino
17	A	3-pyridyl	H	amino
18	B	3-pyridyl	CH_3	2-methoxyphenyl
19	B	3-pyridyl	CH_3	4-chlorophenyl
20	A	3-pyridyl	H	2-(1-imidazoly)ethoxy
21	B	3-pyridyl	H	4-methylphenyl
22	A	2-pyridyl	H	nitro
23	B	3-pyridyl	H	4-cyanophenyl
24	A	3-pyridyl	CH_3	amino
25	A	3-pyridyl	H	chloro
26	A	3-pyridyl	H	hydrogen
27	B	3-pyridyl	H	2-methoxyphenyl
28	B	3-pyridyl	H	2-carboxyphenyl
29	A	3-pyridyl	H	1-imidazolyl
30	B	3-pyridyl	H	2-pyridyl
31	A	3-pyridyl	H	methoxycarbonyl
32	B	3-pyridyl	CH_3	methyl
33	A	3-pyridyl	H	carboxy
34	A	3-pyridyl	H	nitro
35	A	4-chlorophenyl	H	tetrafluoroethoxy
36	A	4-pyridyl	H	nitro
37	A	4-pyridyl	H	tetrafluoroethoxy

Table 1. Phenylaminopyrimidine (PAP) series of compounds prepared for molecular docking [7].

Results

Biosimulated binding between experimental and predicted IC_{50} by AutoDock 3.0 demonstrated positive correlation coefficient (R-value) of 0.677, suggesting relationship of simulated with biochemical enzymatic assay. The PAP group-A with 17 compounds has R-value of 0.625 while, for PAP group-B with 20 compounds has R-value of 0.712. PAP compound-1 (see Table 1), showed lowest binding energy (17.41 kcal.mol^{-1}), implying highest binding affinity interaction to 1FPU protein, in agreement with biochemical assay performed [7]. Interacting model of PAP-1 (STI-571) with 1IEP inactive Abelson kinase showed (Figure 3a) STI-571 as the most selective inhibitor among the 37 PAP compounds, indicated by the number of formed complementary hydrogen bonds. The three N atoms of STI-571 including: N amide linkage, N atom in pyridyl and piperazinyl moieties, particularly act as donor atoms to form a total of four hydrogen bonds with O atoms of Asp381 and Glu286, Met318 and Ile360 residues in 1IEP protein inactive Abl, respectively.

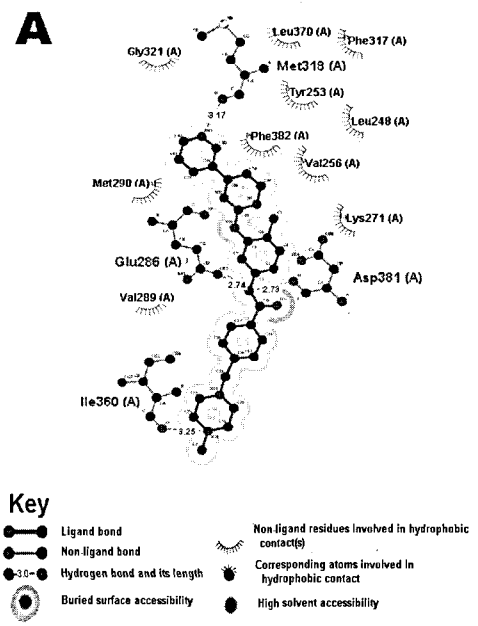


Figure 3a. Modeling of STI-571 with 1IEP protein inactive Abl conformation. 1IEP protein coordinate was taken from PDB [8].

The result of modeling hydrogen bonds of STI-571 with 1IEP inactive Abl form (Figure 3a) is closely related to crystallographic studies, see Figure 3b [9, 10]. STI-571 revealed predicted hydrogen bonds to the activation loop with distances of 3.17Å (Met318), 2.74Å (Glu286), 2.8Å (Asp381) and 3.25Å (Ile360), in close agreement with crystal structure [10] hydrogen bond distances of 2.90Å (Met318), 2.88Å (Thr315), 3.0Å (Glu286), 2.9Å (Asp381), 2.67Å (Ile360) and 3.19Å (His361). Data from Ligplot analysis supported the idea that correlation exists between predicted and experimental studies, significance F = 0.58.

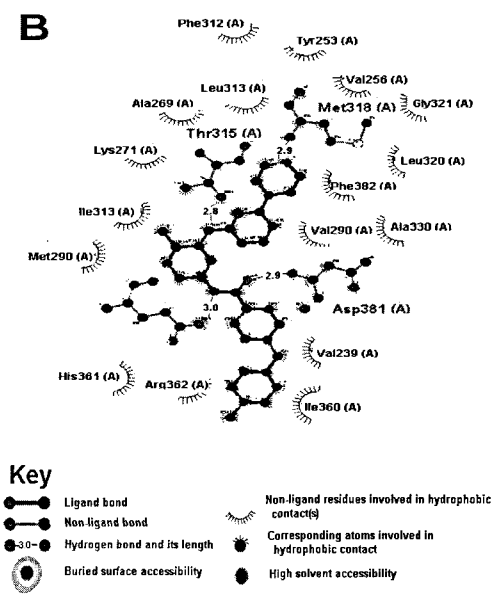


Figure 3b. Crystal structure of bound STI-571 inhibitor with 1IEP inactive Abl.

To determine the juxtaposed atomic positions between experimental and simulated STI-IEP complex, LPC analysis was employed. The crystal structure of bound STI-571 to 1IEP inactive Abl showed [12] inhibitor contact to residues Tyr253A (49.1 $Å^2$), Glu286A (41.1 $Å^2$), Met290A (40.8 $Å^2$), Thr315 (31.1 $Å^2$), Met318A (48.9 $Å^2$) and Asp381A (69.6 $Å^2$). The crystallographic residues that showed similar high contacts were Tyr253A (60.3 $Å^2$), Glu286A (60.7 $Å^2$), Met290A (39.2 $Å^2$), Thr315 (33.9 $Å^2$), Met318A (45.5 $Å^2$) and Asp381A (62.8 $Å^2$).

Discussion

Distinctive PAP series of compounds were able to locate and bind into the active site of Abl kinase complexed structure. The high binding affinity of STI-571 (PAP-1) to inactive Abl kinase among the rest of PAP compounds is in accord with biochemical inhibition assay [7] established. The findings show the efficiency of Lamarckian genetic algorithm in AD3 for identifying the best inhibitor candidate. As observed, there is proximal affinity of ligand to protein in predicted and crystal structures of STI571-1IEP complex. The STI-571 high interatomic affinity to inactive Abl kinase supports the compound's high binding affinity as scored by AD3. Moreover, STI-571 high atomic contacts to inactive Abl illustrated 83% (15 out of 18) matching in reference to X-ray interatomic contact. The reproducible modeling of STI-571 justifies efficiency of LGA in AutoDock3.

However, some of PAP compounds in the series showed not well-correlated (data not shown) predicted inhibition as compared to experimental values of IC_{50} that maybe attributed to the estimated solvation [13] computation inherent in the AD3 program. Apparently, a serious limitation [14] in many existing scoring functions, including AD3, is the tendency to use solvent models in rough estimations.

AD3 has high efficiency in locating the active site and discriminating highly weak from strong ligand binders. However, caution must be taken into consideration in distinguishing similar affinities of different inhibitors contributed by solvation effect limitation.

Acknowledgement

This research was supported by scholarship grant from the Philippine Council for the Advancement of Science and Technology Research and Development. In addition, the Center for Integrative and Development Studies, UP-Diliman awarded research financial assistance.

Gratitude to Prof. Arthur Olson from The Scripps Research Institute for allowing the free use of AutoDock3. Dr. Roman Laskowski from European Bioinformatics Institute provided the decryption key free of charge. Dr. Sobolev from Weizmann Institute of Science created an on-line easy to use tool for the analysis of ligand-protein contacts.

References

[1] Morris, G., Goodsell, D., Halliday, R., Hart, W., Belew, R. and Olson, A. Automated Docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal Computational Chemistry,* 19(14): 1639-1662 (1998).

[2] Drexler, H., Macleod, R. and Uphoff, C. Leukemia Cell Lines: in vitro models for the study of Philadelphia chromosome-positive leukemia. *Leukemia Research,* 23(3): 207-215 (1999).

[3] von Bubnoff, Scheneller, Peschel and Duyster. Bcr-Abl genes mutations in relation to clinical resistance of Philadelphia chromosome-positive leukemia to STI-571: a prospective study. *Lancet,* 359:487-491 (2002).

[4] Schindler, T., Bornmann, W., Pellicina, P., Todd Miller, W., Clarkson, B. and Kuriyan, J. Structural mechanism for STI-571 inhibition of Abelson tyrosine kinase. *Science,* 289: 1938-1942 (2000).

[5] Bergan, R., Waggle, D., Carter, S., Horak, I., Slichenmyer, W. and Meyers, M. Tyrosine kinase inhibitors and signal transduction modulators: rationale and current status as chemopreventive agents for prostate cancer. *Urology,* 57 (4A Suppl S): 77-80 (2001).

[6] Wisniewski, D., Lambek, C., Liu, C., Strife, A., Veach, D., Nagar, B., Young, M., Schindler, T., Bornmann, W., Bertilion, J. and Kuriyan, J. Characterization of potent inhibitors of the Bcr-Abl and the c-kit receptor tyrosine kinase. *Cancer Research,* 62:4244-4255 (2002).

[7] Zimmermann, J., Buchdunger, E., Mett, H., Meyer, T. and Lydon, N. Potent and selective inhibitors of the Abl kinase phenylaminopyrimidine (PAP) derivatives. *Bioorganic and Medicinal Chemistry Letters,* 7(2): 187-192 (1997).

[8] Protein data bank: structure explorer-1IEP, download/display file.
URL:http://www.rcsb.org/pdb/cgi/explore.cgi?job=download&pdbId=1IEP&page=&pid=978210582383376.

[9] Protein data bank: summary PDB structure-1IEP, ligplot interaction.
URL:http://www.biochem.ucl.ac.uk/bsm/pdbsum/1iep/ligplot01.html.

[10] Nagar, B., Bornmann, W., Pellicina, P., Schindler, T., Veach, D., Todd Miller, W., Clarkson, B. and Kuriyan, J. Crystal structures of the Abl kinase domain of c-Abl in complex with small molecule inhibitors PD17119 and imatinib (STI-571). *Cancer Research,* 62: 4236-4243 (2002).

[11] Haruyama, H., Ohkuma, V., Nagaki, H., Ogita, T., Tamaki, K. and Kinoshita, T. Matlystatins, new inhibitors of type IV collagenases from actinomadura atramentaria III. Structure elucidation of matlystatin A to F. *Journal of Antibiotic (Tokyo)* 47(12): 1473-1480 (1994).

[12] Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. and Edelman M. Automated analysis of interatomic contacts in proteins. *Bioinformatics,* 15: 327-332 (1999).

[13] Csaba, H. and van der Spoel, D. Efficient docking of peptides to proteins without prior knowledge of the binding site. *Protein Science,* 11: 1729-1737 (2002).

[14] Akaho, E., Morris, G., Goodsell, D., Wong, D. and Olson, A. A study on docking mode of HIV protease and their inhibitors. *Journal of Chemical Software,* 7(3): 103-114 (2001).