American Medical Journal 1 (2): 126-132, 2010
ISSN 1949-0070

Analysis of Resistance to Human Immunodeficiency Virus Protease Inhibitors Using Molecular Mechanics and Machine Learning Strategies

[1]A. Harishchander, [1]D. Alex Anand and [2]Sanjib Senapati
[1]Department of Bioinformatics, Sathyabama University, Chennai, India
[2]Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India

Abstract: Problem statement: Drug resistance is the most important factor, which influences the failure of current HIV therapies. So, the ability to predict the drug resistance of HIV-1 protease mutants will be useful in developing more effective and longer lasting treatment regimens. **Approach:** Drug resistance of HIV-1 protease is predicted with two current protease inhibitors (Indinavir and Saquinavir). The problem was approached from two perspectives. First, structural features of the HIV protease with inhibitor complex were constructed. Next, a classifier was constructed based on the patterns of various drug resistant mutants. In first stage SPDB viewer (for making mutations) and INSIGHT II (for analyzing binding energies and hydrogen bond contact with the inhibitor and the binding site) software's were used for structural property analysis. In the second stage a supervised learning linear Classifier (SVM-LIB) in DTREG tool has been used to analyze the Resistant and susceptible patterns. Finally Genetic Algorithm in Matlab tool has been used for Optimization. **Results:** Structural data mining performed linear SVM model gives "93% accuracy" in initial screening of pattern sets HIV1 protease (wild type and mutants) of sub type B against the inhibitors Indinavir and Saquinavir. Genetic algorithm gives "80% Accuracy" for Indinavir and "60% Accuracy" for Saquinavir. **Conclusion:** Geno2pheno software uses machine learning analysis for subtypes of HIV with proper inhibitory values. If Molecular Mechanics is followed by Machine Learning with appropriate Inhibitory or effective concentration analysis, the validation of Genotyping will be more accurate than initial Geno2Pheno analysis. In future even the dynamics of the molecule will be analyzed with molecular mechanics and machine learning principles for various mutations of all FDA approved protease Inhibitors within the individual complex with the protease.

Key words: Binding energies, protease, hydrogen bonding, linear SVM model, Matlab tool and genetic algorithms

INTRODUCTION

HIV has pol gene to code 3 enzymes namely Protease, Reverse Transcriptase and Integrase. HIV protease is an aspartyl D protease with 99 amino acids (Basu *et al*., 2004a; 2004b). The three dimensional structure of the HIV protease contains primarily β-sheet, turn and extended structural elements. The structure is unusual in that the dimer has only one active site. Each monomer contributes one of the two aspartyl residues within the Asp-Thr-Gly sequences of the active site. The HIV protease active site is located in a cleft into which the polypeptide to be cleaved is positioned (Ishima, 1999; Pattabiraman, 1999). Substrate or inhibitor binding to the protease induces a large conformational change. The flaps of the protease move as much as 15°A when the ligand is bound (Louis, 2001). First, the mechanisms are examined that allow the HIV virus to develop drug resistance to the FDA-approved protease inhibitor Indinavir. The structural changes that characterize drug resistant protease mutants are studied in order to understand the effect that various structural changes have upon drug resistance (Winters, 2000; Xie, 1999; Louis, 2001). The results suggest that the drug resistance phenomenon is associated with a loss of contacts between the drug and the target viral enzyme. A further observation is that different point mutations may lead to similar structural changes in the active site (Mahalingam *et al*., 1999). This study also investigated the resistance of HIV protease mutants to Saquinavir (another FDA approved protease inhibitor). No attempt is made to understand the mechanism or reasons why certain mutations are not resistant to Saquinavir, but rather to predict such

Corresponding Author: D. Alex Anand, Department of Bioinformatics, Sathyabama University, Chennai, India
Tel: +91-44-24501644/+91-44-24962208/+91-98402-35781/Ext: 5213 Fax: +91-44-24501644-2344

resistance based solely on the amino acid sequence of HIV protease mutants. A small number of these mutants have reported Saquinavir IC_{90} values, which were used to classify the resistance of the mutants tested (Draghici and Potter, 2003). Once a treatment failure has been detected, the usual measure is to change the treatment and attack the virus with a different combination of drugs. There are two major problems here (Draghici and Potter, 2003). First, the number of FDA-approved drugs is limited and therefore the number of effective combinations of drugs is also limited. It is conceivable that a viral quasi-species may become resistant to all known drugs, thus rendering the treatment ineffective. A second problem is that of Cross-resistance, which further reduces the number of effective combination therapies (Draghici and Potter, 2003). So Molecular Mechanics followed by machine learning analysis is performed to analyze and understand the complete resistance of inhibitors against HIV protease.

Trying to relate the structure of the virus to drug resistance by understanding the general relationship between the structure and function of the HIV virus. Various researches studying the HIV protease have focused on the flaps and dimer-interface flexibility (Ishima, 1999). Molecular surface analysis (Pattabiraman, 1999) and the auto-processing of the HIV-1 protease (Louis, 2001). The drug resistance problem was also studied in the context of various Mutations (Mahalingam *et al*., 1999) and the folded monomer of HIV protease (Louis, 2001).The HIV protease has become a prime target in drug design since the discovery that the chemical inhibition or the mutational inactivation of the enzyme generates noninfectious viral particles. Furthermore, the protease gene has a relatively small coding region and the enzyme and stringent substrate specificity (Draghici and Potter, 2003).

First, the mechanisms are examined that allow the HIV virus to develop drug resistance to the FDA-approved protease inhibitor Indinavir (Draghici and Potter, 2003). The structural changes that characterize drug resistant protease mutants are studied in order to understand the effect that various structural changes have upon drug resistance. The results suggest that the drug resistance phenomenon is associated with a loss of contacts between the drug and the target viral enzyme (Louis, 2001; Mahalingam *et al*., 1999; Winters *et al*., 1998). Further observation is that different point mutations may lead to similar structural changes in the active site (Ishima, 1999; Pattabiraman, 1999; Durant, 1999). This study also investigated the resistance of HIV protease mutants to Seqinavir (another FDA approved protease inhibitor). No attempt is made to understand the mechanism or reasons why certain mutations are not resistant to Saquinavir, but rather to predict such resistance based solely on the amino acid sequence of HIV protease mutants. A small number of these mutants have reported Saquinavir IC_{90} values, which were used to classify the resistance of the mutants tested. Once a treatment failure has been detected, the usual measure is to change the treatment and attack the virus with a different combination of drugs.

IC_{90} values: The IC_{90} is the result of a "Phenotypic" test and represents the amount of drug necessary to reduce viral replication by 90% (Draghici and Potter, 2003). The ratio between the IC_{90} value of a specific mutant and the IC_{90} value of the wild type (the virus in its non-mutated form) is the fold resistance. IC_{90} value refers to 90% inhibition concentration shown by the drug towards viral strain (Winters *et al*., 1998) (Fig. 1 and 2).

ID	Mutation	IC_{90} (uM)	Fold resistance
-	Wild Type	0.12	1
1	L10I	4.08	34
2	G48V	9.12	76
3	I54V	NR	NR
4	G73S	NR	NR
5	V82A	NR	NR
6	I84V	2.28	19
22	L10I K14R N37D M46I F53L A71V G73S V77I L90M	1.57	13.08
23	L10I E35D M36I R41K I62V L63P A71V G73S I84V L90M I93L	9.14	76.17
24	L10I I15V M36I G48V I54V I62V V82A	0.66	5.5
25	L10I I15V M36I G48V I54V I62V	0.57	4.75
26	K14R I15V N37D F53L A71V G73S L90M	1.09	9.08
27	K14E M36V G48V L63P A71V T74S V82A	0.65	5.42
28	I15V R41K L63P A71T G73S L90M	0.72	6.0
29	G48V L63P T74A	0.35	2.92
30	K20I M36I L63P A71T G73S L90M	0.43	3.58
31	L10I E35D R41K I62V L63P A71V G73S I84V L90M I93L	0.28	2.33
32	K14R R41K L63P V77I L90M I93L	0.41	3.42
33	L10I K20M L63P A71T V77I L90M I93L	0.32	2.67
34	N37D R57K D60E L63P A71V G73S L90M I93L	0.22	1.83
35	I15V D30N E35D M36I R41K L63P	0.08	0.67
36	L63P T74S L90M	NR	NR
37	L63P L90M	NR	NR
38	K14R R41K L63P V77I I93L	0.16	1.33
39	L10V I62V G73S L90M	0.05	0.42
40	L63P T74A V77I	NR	NR
41	L63P L90M	0.07	0.58
42	N37D L63P A71V G73S L90M I93L	0.15	1.25
43	L10I L63P A71T V77I I93L	0.11	0.92
44	I15V E35D R41K L63P	0.08	0.67
45	K14R/K L63P I93I/L	NR	NR
46	K14E L63P A71V	NR	NR
47	I15V	NR	NR
48	L63P	0.02	0.17
49	L10I L63T A71T	NR	NR
50	L63P A71V L90M	NR	NR
51	L63A	NR	NR
52	G48V I54V L90M	NR	NR
53	G48V I84V L90M	NR	NR

Fig. 1: IC_{90} values of Indinavir against sub type B strain

Mutation	IC_{90} (uM)	Fold resistance
Wild Type	0.03	1
L10I K14R N37D M46I F53L A71V G73S V77I L90M	8.08	269
L10I E35D M36I R41K I62V L63P A71V G73S I84V L90M I93L	6.00	200
L10I I15V M36I G48V I54V I62V V82A	1.18	39
L10I I15V M36I G48V I54V I62V	0.92	30.67
K14R I15V N37D F53L A71V G73S L90M	0.58	19
K14E M36V G48V L63P A71V T74S V82A	0.58	19
I15V R41K L63P A71T G73S L90M	0.37	12
G48V L63P T74A	0.80	27
K20I M36I L63P A71T G73S L90M	0.42	14
L10I E35D R41K I62V L63P A71V G73S I84V L90M I93L	0.34	13
K14R R41K L63P V77I L90M I93L	0.21	7
L10I K20M L63P A71T V77I L90M I93L	0.20	7
N37D R57K D60E L63P A71V G73S L90M I93L	0.20	7
I15V D30N E35D M36I R41K L63P	0.03	1
L63P T74S L90M	0.09	3
L63P L90M	0.08	3
K14R R41K L63P V77I I93L	0.07	2
L10V I62V G73S L90M	0.07	2
L63P T74A V77I	0.07	2
L63P L90M	0.06	2
N37D L63P A71V G73S L90M I93L	0.06	2
L10I L63P A71T V77I I93L	0.06	2
I15V E35D R41K L63P	0.06	2
K14R/K L63P I93I/L	0.06	2
K14E L63P A71V	0.06	2
I15V	0.04	1
L63P	0.05	2
L10I L63T A71T	0.02	1
L63P A71V L90M	0.02	1
L63A	0.01	0.33
G48V I54V L90M (Schinazi *et al.*, 1999)	1.50	50
G48V I84V L90M (Schinazi *et al.*, 1999)	0.90	30

Fig. 2: IC_{90} values of Saquinavir against sub type B strain

Note:

- The fold resistance was calculated as a ratio between the IC_{90} value of the mutant and the IC_{90} value of the wild type. All mutations were obtained from Winters *et al*. (1998), except as noted
- Resistance values were not available for 14 out of the total of 38 mutants with resistance and 14 with NR. (NR = No Resistance Reported. Patterns 1-6, 52 and 53 are from Schinazi *et al*. (1999). Patterns 7-21 are also from Schinazi *et al*. (1999) and Winters *et al*. (1998), but are not included because they are redundant with patterns from Winters *et al*. (1998)

MATERIALS AND METHODS

Structure-based data mining-PROTOCOL:

- Construct mutant genotypes and produce 3D structures using SPDB viewer (i.e., for making mutations (Fig. 3))

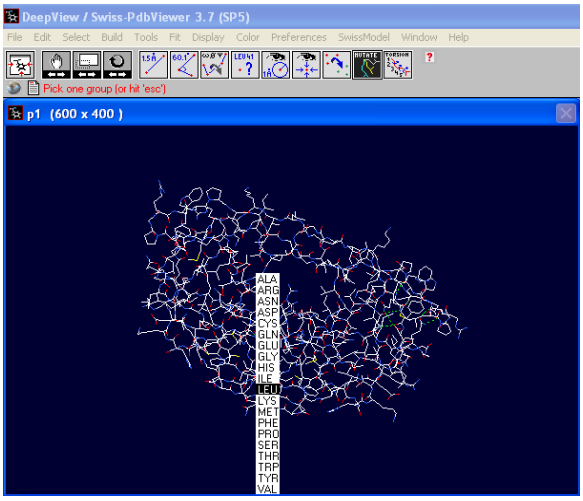

Fig. 3: Performing mutations in SPDB viewer

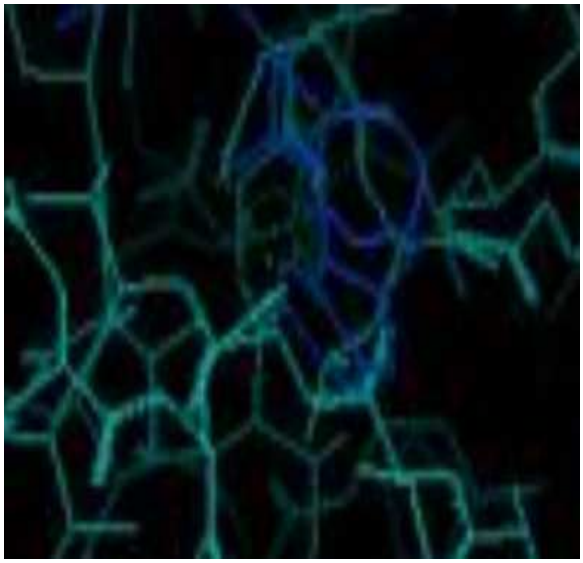

Fig. 4: Analyzing hydrogen bonds between Saquinavir and wild type HIV protease using INSIGHT-II. (Here there are 2 hydrogen bonds)

- Use INSIGHT II to analyze the 3D structures and produce a list of contacts between the mutant proteases and protease inhibitor (i.e., analyzing hydrogen bonding and binding energies (Fig. 4-7))
- Preprocess the contact information (input reduction normalization)
- Construct and train the patterns using Support Vector Machine (SVM) to categorize Mutant resistance to the protease inhibitor using DTREG tool

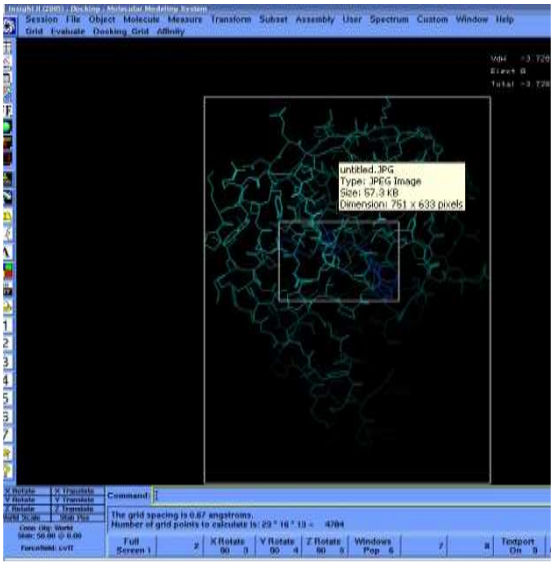

Fig. 5: Analyzing binding energy between Saquinavir and wild type HIV protease using INSIGHT-II. (Here the binding energy is -3.728 Kcal mole^{-1})

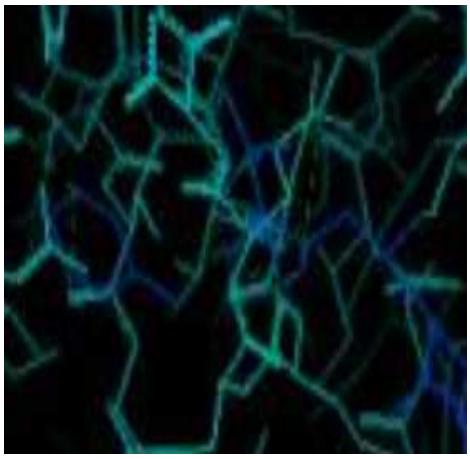

Fig. 6: Analyzing hydrogen bonds between Indinavir and wild type HIV protease using INSIGHT-II. (Here there are 4 hydrogen bonds)

- Test the SVM and analyze its performance and further perform virtual screening with genetic algorithm using Matlab software (Fig. 8 and 9)

Advantages of structure base data mining:

- Easy to understand (Meta class construction and Prediction mapping)
- More parameters are not required for modeling.
- More reliable confinement of appropriate accuracy

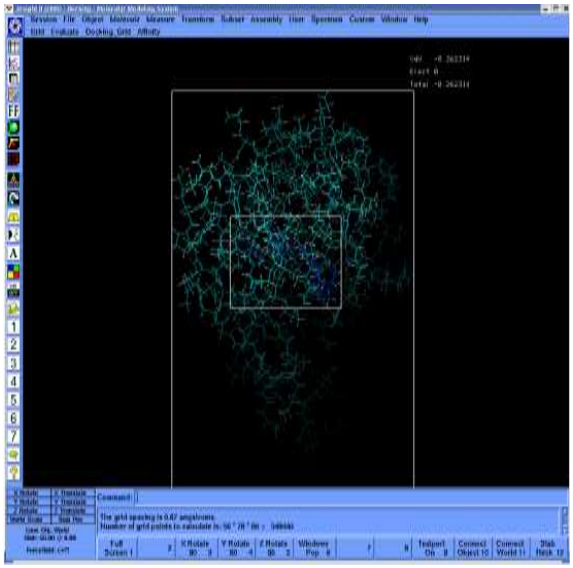

Fig. 7: Analyzing binding energy between Indinavir and wild type HIV protease using INSIGHT-II. (Here the binding energy is-0.262 Kcal mole^{-1})

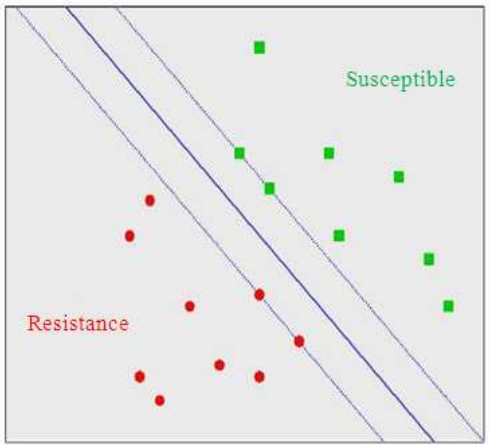

Fig. 8: Output of linear-SVM predictions on Indinavir (34 patterns), optimized to linear graph from DTREG tool

- No need for the Iteration of Learning rates with neighborhood method to calculate score:

$$\text{Score} = \text{Coverage} \times \text{Accuracy} \times 100$$

- Implementation of algorithm is much easier when compare to neighborhood calculations

Graphical representation of mutant construction followed by molecular analysis:

Initial screening meteorology (Indinavir): Done by support vector machine, in training higher weightage is

given for binding energy (5) and plotted on Y axis and lower weightage is given for H bonding (1) and plotted on X-axis.

Initial screening meteorology (Seqinavir): Done by support vector machine, in training higher weightage is given for binding energy (3) and plotted on Y axis and lower weightage is given for H bonding (1) and plotted on X-axis.

Calculation of accuracy in linear SVM model (Indinavir) (Table 1): Accuracy calculation (true-positive state):

(Resistant Patterns × Sensitivity)/(Susceptible Patterns) × (1-Specificity) × 100 = (7×0.4)/[6×(1-0.5)]×100 = 93%

Note:

- Sensitivity is 0.4, since 4 fold cross validation has been performed for 10 iterations with 1 threshold resistant pattern

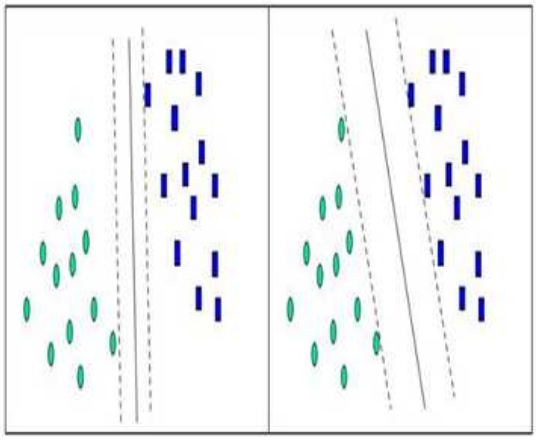

Fig. 9: Output of linear-SVM predictions on Saquinavir (31 patterns), optimized to comparative marginal graph from DTREG tool

Table 1: Classification of Indinavir resistance (initial screening)

Class	Patterns	Fold resistance
High resistance	6, 1, 22	19, 34, 13.08
Low resistance	24, 27, 28, 20	5.5, 4.2, 6.0, 3.58
High suceptability	34, 42, 33	1.83, 1.25, 2.67
Low suceptability	38, 31, 25	1.33, 2.33, 4.75

Table 2: Classification of Saquinavir resistance (initial screening)

Class	Patterns	Fold resistance
High resistance	4, 8, 2	30.67, 27, 200
Low resistance	1, 3	269, 30.67
High suceptability	16, 21, 5	27, 1, 3
Low suceptability	9, 24	7, 1

- Specificity is 0.5, since (13 points+4 support points) are classified in a linear graph for 10 iterations

Calculation of accuracy in Linear SVM model (Saquinavir) (Table 2): Accuracy calculation (true-positive state):

[(Resistant Patterns × Sensitivity)/(Susceptible Patterns) × (1-Specificity)] × 100 = (5×0.8)/[5×(1-0.14)] ×100 = 93%

Note:

- Sensitivity is 0.8, since 4 folds cross validation has been performed for 10 iterations for 10 points with 2 thresholds training resistant patterns
- Specificity is 0.14, since points (10 points+2 support points) are classified in a marginal graph for 10 iterations

Optimized resistance (Indinavir): Done by fitness principles in genetic algorithm using Matlab tool (Fig. 10).

Optimized resistance (Saquinavir): Done by fitness principles in genetic algorithm using Matlab tool (Fig. 11).

Calculation of accuracy in genetic algorithm for Indinavir (optimized resistance) (Table 3): Accuracy calculation (positive fitness):

[(Best Fitted Values) × (cross validated profile) × (Non Fitted Values) × (Eigen Test Fitness)]/[(Number of Generation) × (Mean Value of training set)] × 100 = [(3×2×6×10)/(100×4.5)] × 100 = 80%

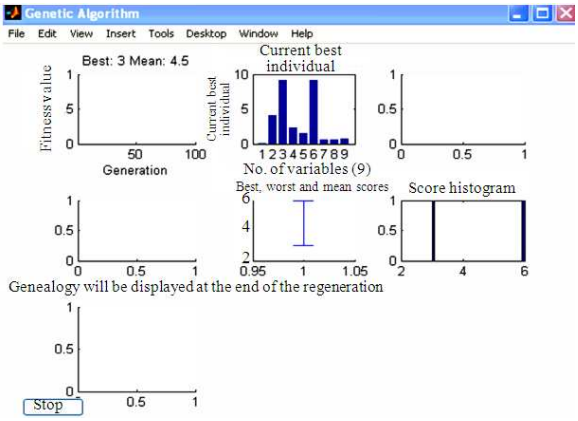

Fig. 10: Output of genetic algorithms predictions on Indinavir (9 patterns)

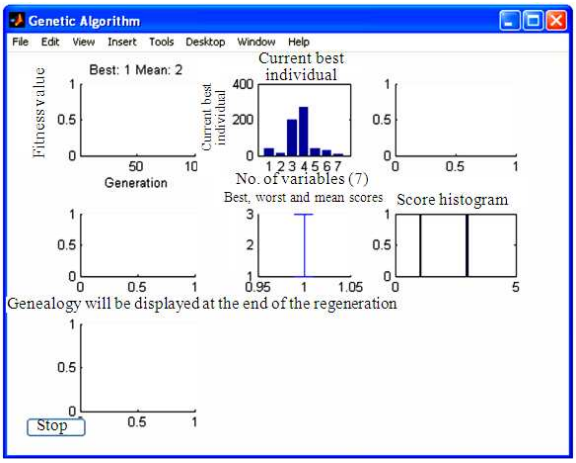

Fig. 11: Output of genetic algorithms predictions on Saquinavir (9 patterns)

Table 3: Classification of complete resistance of Indinavir (optimized resistance)

Class	Patterns	Fold resistance
High resistance	24, 28, 1	5.5, 6.0, 19
Low resistance	22	13.08

Table 4: Classification of complete resistance of Saquinavir (optimized resistance)

Class	Patterns	Fold resistance
High resistance	2, 1	200, 269
Low resistance	4, 3	39, 39.67

Note:

Eigen test fitness is 10 for 100 generations.

Calculation of accuracy in genetic algorithm for Saquinavir (optimized resistance) (Table 4): Accuracy calculation (positive fitness):

[(Best Fitted Values) × (cross validated profile) × (Non Fitted Values) × (Eigen Test Fitness)]/[(Number of Generation) × (Mean Value of training set)] ×100 = [(1×2×6×10)/(100×2)]×100 = 60%

Note:

Eigen test fitness is 10 for 100 generations.

Implementation of usage:

- To develop drug resistance for the failure of current HIV therapies by HAART methods
- The ability to predict the drug resistance will be useful in developing long lasting treatment regimens

Logical implementation (hypothesis): If the number of resistant patterns of the drug are greater than the Number of Susceptible patterns of the drug, excluding the support vectors and unbound patterns, then the drug molecule can be taken for further analysis of treatment regimen. If needed Genetic algorithms can be used to analyze the complete resistance of the drug towards certain patterns.

RESULTS

Comprehensive analysis of hydrogen bonding and binding Energies are taken for various mutants and trained in two classifying methods namely Linear SVM method and genetic algorithms structural data mining performed Linear SVM model gives “93% accuracy” in initial screening of pattern sets HIV1 protease (wild type and Mutants) of sub type B against the inhibitors Indinavir and Saquinavir (Fig. 8 and 9). Genetic algorithm gives “80% accuracy” for Indinavir and “60% accuracy” for Saquinavir.

DISCUSSION

Results were obtained from software’s SPDB Viewer (Fig. 3), INSIGHT II (Silicon graphics Machine in Unix Platform using Grid based conjugate gradient algorithm), DTREG tool Optimization and Matlab optimization. Among various sets of descriptors only hydrogen bonding (Fig. 4 and 6) and integrated force field energy parameters (binding energy (Fig. 5 and 7)) are assumed to be the fit parameters for analyzing the stability of the molecule, i.e., molecular mechanics. So by accurate predictions of these properties from insight-II software gives us the details of molecular stability.

CONCLUSION

Structural data mining analysis shows that the screening method is Linear SVM (Acevedo-Rodríguez *et al*., 2009) model and the confined method is Genetic algorithm. Genetic algorithm is performed using Matlab. This meteorology has to be implemented for in-depth Geno2Pheno analysis, since a better data tree is constructed. So Geno2Pheno analysis can be achieved by both molecular mechanics and machine learning strategies along with proper values of Inhibitory or effective concentrations.Geno2pheno software uses machine learning analysis for subtypes of HIV with proper inhibitory values. If molecular mechanics is followed by machine learning with appropriate Inhibitory or effective concentration. Then the validation of

Genotyping will be more accurate than initial Geno2Pheno analysis. So all these combined confined parameters of Molecular network will make sure to achieve better accuracy for Geno2Pheno analysis.

ACKNOWLEDGEMENT

It is a great pleasure to thank many people who made this study possible. Our sincere thanks to our chancellor Dr. Jeepiar, for allowing us to complete this study at the prestigious Sathyabama University. We would like to express my deep and sincere gratitude to our Directors Mr. Marie Johnson and Mrs. Mariazeena Johnson, for all the funding provided and also to present a part of this study at the WABT conference at Italy. We dedicate this study to Dr. Indrapal Singh Aidhen, Ph.D., Professor, Chemistry Department, Indian Institute of Technology Madras for encouraging Harish towards this area Molecular Studies Finally warm word of gratitude to Harish's parents for supporting this academic endeavor in every way.

REFERENCES

Acevedo-Rodríguez, J., S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Siegmann and F. López-Ferreras, 2009. Computational load reduction in decision functions using support vector machines. Single Process., 89: 2066-2071. http://agamenon.tsc.uah.es/Investigacion/gram/publications/sp09-acevedo.pdf

Basu, S., M. Bilenko and R.J. Mooney, 2004a. A probabilistic framework for semi-supervised clustering. Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 22-25, ACM Press, Seattle, WA., USA., pp: 59-68. http://portal.acm.org/citation.cfm?id=1014062

Basu, S., M. Bilenko and R.J. Mooney, 2004b. Semi-supervised clustering for intelligent user management. http://domino.watson.ibm.com/acas/w3www_acas.nsf/images/projects_03.04/$FILE/mooney_paper.pdf

Draghici, S. and R.B. Potter, 2003. Predicting HIV drug resistance with neural networks. Bioinformatics, 19: 98-107. http://cat.inist.fr/?aModele=afficheN&cpsidt=14491440

Durant, J., 1999. Drug-resistance genotyping in HIV-1 therapy: The VIRAD APT randomized controlled trial. Lancet, 353: 2195-2199. http://linkinghub.elsevier.com/retrieve/pii/S0140673698122912

Ishima, R., 1999. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease and their implications for function. Struct. Fold Des., 7: 1047-1055. http://www.ncbi.nlm.nih.gov/pubmed/10508781

Louis, J.M., 2001. Folded monomer of HIV-1 protease. 276: 49110-49116. http://www.jbc.org/content/276/52/49110.full.pdf+html

Mahalingam, B., J.M. Louis, C.C. Reed, J.M. Adomat and J. Krouse *et al*., 1999. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. Eur. J. Biochem., 263: 238-245. http://www3.interscience.wiley.com/cgi-bin/fulltext/121360605/PDFSTART

Pattabiraman, N., 1999. Occluded molecular surface analysis of ligand-macromolecule contacts: Applications to HIV-1 Protease-inhibitor complexes. J. Med. Chem., 42: 3821-3824. http://pubs.acs.org/doi/pdf/10.1021/jm980512c?cookieSet=1

Schinazi, R.F., B.A. Larder and J.W. Mellors, 1999. Mutations in retroviral genes associated with drug resistance. Int Antivir. News, 4: 46-49. http://direct.bl.uk/bld/PlaceOrder.do?UIN=062729880&ETOC=RN&from=searchengine

Winters, M.A., J.M. Schapiro, J. Lawrence and T.C. Merigan, 1998. Human immunodeficiency virus type 1 protease genotypes and *in vitro* protease inhibitor susceptibilities of isolates from individuals who were switched to other protease inhibitors after long-term Saquinavir treatment. J. Virol., 72: 5303-5306. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC110130/

Winters, M.A., 2000. Frequency of antiretroviral drug resistance mutations in HIV-1 strains. Antivir. Ther., 5: 57-63. http://www.cpcra.org/docs/pubs/papers/CPCRA046-C.pdf

Xie, D., 1999. Drug resistance mutations can affect dimer stability of HIV-1 protease at neutral pH. Protein Sci., 8: 1702-1707. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2144419/pdf/10452615.pdf