iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 6 **Issue:** IV **Month of publication:** April 2018

DOI: http://doi.org/10.22214/ijraset.2018.4306

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 6.887
Volume 6 Issue IV, April 2018- Available at www.ijraset.com

Proteochemometrics: A Bio-Statistical Approach to Drug Design

Sandeep Singh[1], Akanksha Rajpoot[2], Kshipra Deshmukh[3]
[1, 2, 3,] *Biotechnology Department, MGM's College of Engineering & Technology, Kamothe, Navi Mumbai*

Abstract: ***Integrase is an enzyme that is produced by HIV viruses or rather all the retroviruses that enables its genetic material to be integrated into the DNA of the infected cell. It is also produced by the viruses containing double stranded DNAs for the same purpose. Inhibition of this enzyme can be a major success in the treatment of HIV (AIDS). The major obstacle in the treatment of HIV is the ability of the virus to mutate rapidly into drug resistant mutants. For this reason a new a QSAR-like approach, Proteochemometrics, was introduced a few years ago, which remedies these problems. It has the capacity to generalize between targets potentially, to the extent that it could provide proteome-wide drug screening over all of the existing targets. It is an approach which overcomes the drawbacks of QSAR and HTS. By adopting Proteochemometrics for modeling of a drug as HIV1 Integrase inhibitor, it was seen that the R^2 calculated for the principal components of receptors was 0.80 whereas the R^2 calculated for the 'crossterms' of principal components of receptors and ligands, was 0.85 which indicates that proteochemometric modeling is a better approach for designing new drugs, as it increases the robustness of the models for drug development.***

Keywords: ***Proteochemometrics, Integrase, HIV Inhibitors***

I. INTRODUCTION

A. HIV (Human Immunodeficiency Virus)

Since HIV is a virus, it cannot grow or reproduce on their own, they need to infect the cells of a living organism in order to replicate. The human immune system usually finds and kills viruses fairly quickly, but HIV attacks the immune system itself – the very thing that would normally get rid of a virus. With around 2.7 million people becoming infected with HIV in 2008, there are now an estimated 33 million people around the world who are living with HIV, including millions who have developed AIDS [1].

B. Drugs Targeting Integrase

The compounds whose antiviral activity was analyzed against integrase include the following: L870, S1360, 5CITEP, GS9137, GSK364735, MK2048 and MK0518 or Raltegravir [2, 3]. Raltegravir is the only drug existing for therapeutic use. Rest all the compounds are currently under different phases of trials.

C. The Proteochemometric Approach

This is an analytical tool quantifying the activity of a drug with respect to its target. The method exploits affinity data for a series of organic chemical compounds binding to wild-type and artificially mutated receptors. The receptor sequences and compounds are assigned predictor variables that are correlated to the measured pharmacological activities using partial least-squares projections to latent structures [4]. The predictor variables consist of one descriptor block derived from the chemical properties of the receptors' primary amino acid sequences and another descriptor block derived from the chemical properties of the organic compounds. The cross-terms generated from the two descriptor blocks are also derived. Moreover, models also give quantitative information about the interactions of the amino acids of the receptors with the ligands, thereby giving an insight into the molecular mechanisms involved in ligand binding [5].

In our study we have taken HIV1-Integrase enzyme as the target of which there are 30 mutants. These mutations in the enzymes are caused due to the administration of different regimens of integrase inhibitors to the infected people. Different descriptors of these mutated sequences and the inhibitors were calculated and then correlated. Regression analysis was done to find the R^2 and predict the model.

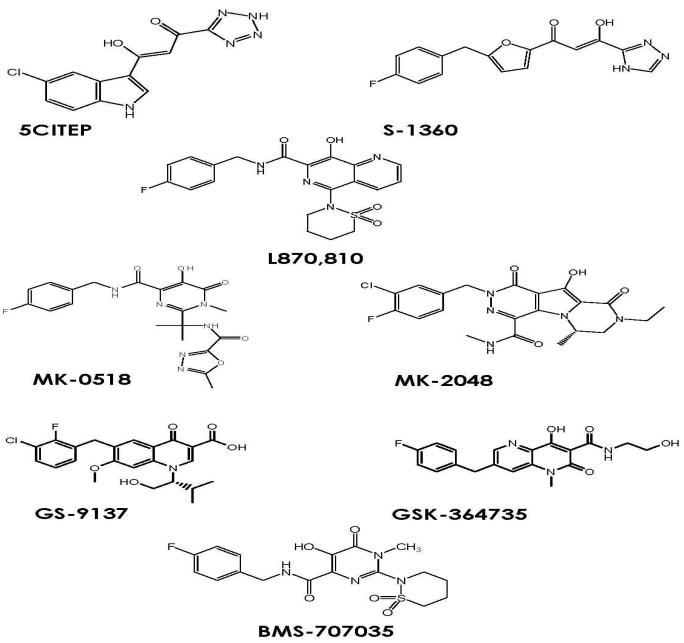
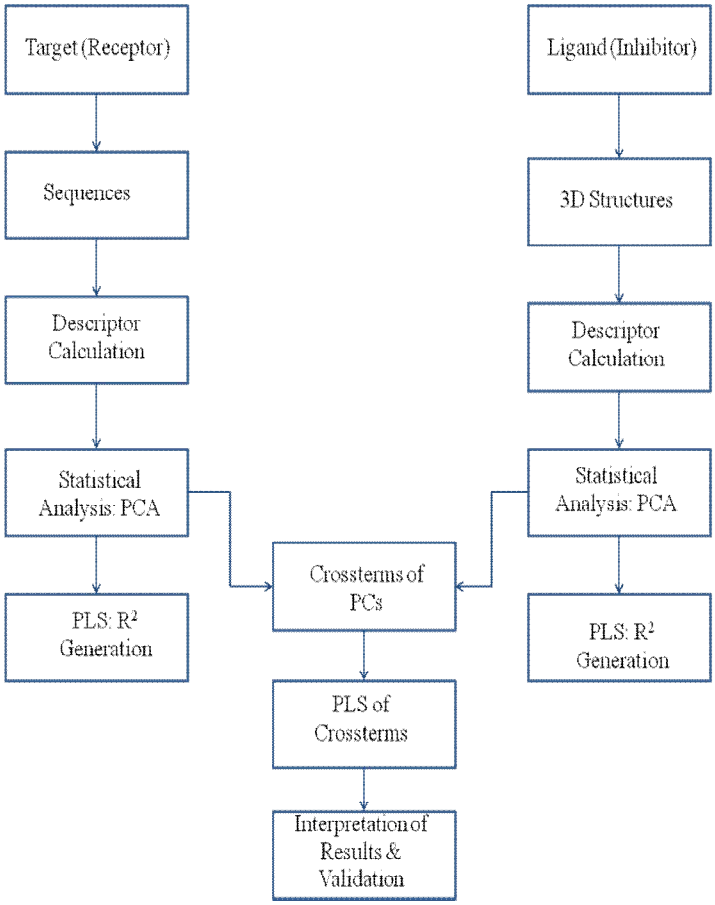


Fig. 1: The structure of diketo acid-based HIV-1 integrase inhibitors.

II. METHODOLOGY



A. Acquiring The Target Sequence: Hiv

Integrase was decided upon as the protein (target) of interest due to its importance. Integrase protein sequence mutates after the administration of drug Raltegravir, along with certain other factors which influence the behavior of Integrase priotein (HIV1 IN). Stanford HIV Drug Resistance Database was accessed and 30 instances of mutated Integrase protein, due to the effect of Raltegravir, were isolated.

B. Acquiring Ligand Structure

Raltegravir (a.k.a. Isentress™) became the first FDA approved IN inhibitor in October is currently being administered as a new addition to HAART (Highly Active Antiretroviral Therapy) regimens. [10]

C. Calculating the descriptors

Descriptors are numerical values. These values were calculated using web servers Profeat (for protein) and Molfeat (for ligand). These numerical values lie in a certain range pertaining to that protein or ligand. If the values are more towards upper limit of the range, the descriptor is considered to be a crucial one for that molecule. [11]

D. Principal Component Analysis (PCA)

Fifty-five descriptors of thirty protein and twenty-five descriptors of seven ligands formed a big mass of data. Handling such a huge data may have caused human error and also a lot of confusion. Hence PCA was done using Unscrambler.

E. Numerical values indicating the activity

of the ligand were recorded from literature (12 through to 14). These values were IC50 or inhibitor concentrations are the values that indicate the concentration of the inhibitor at which fifty percent population of virus was inhibited.

F. Cross-terms

are the main feature of proteochemometrics approach. These values represent the interaction space being the space spanned by the non-covalent interactions between all targets and all ligands. Cross-terms of principal components (PCs) of ligand and proteins were calculated manually using MS Excel formula.

G. Multiple Linear Regressions (MLR)

MLR was calculated using the software Statgraphics. On one hand, in three columns, logarithmic values of the ninety descriptors for proteins were input which were the 'Independent Variables'. On the other hand, in the fourth column, logarithmic values of the activity of the ligand were input. This formed the 'Dependant Variable'. MLR was calculated and R^2 was obtained.

III. RESULTS AND DISCUSSION

The results from the above methods were taken and then they were aligned. As mentioned above the mutated sequence of the HIV1 Integrase from the patients were taken from the Stanford HIV Drug Resistance Database. The results of each of the patients were in the tabular format as given below:

A. Integrase Sequence

INIMajorDRMs	INIMinorDRMs	Polys	UnusualMuts
G140A, Q148R, E157Q	V165IV	K14R, S24G, D25E, L45I, I113V, T124N, T125A, I220L, Y227F, L234V, 256E	K160S

B. Integrase

Table 1: Descriptor values of sequence 1 to sequence 5, there were many descriptors which were given as output from Profeat Server out of which 10 descriptors were chosen for the analysis.

	Seq 1	Seq 2	Seq 3	Seq 4	Seq 5
D1	0.064711	0.065766	0.058243	0.058552	0.058552
D2	0.122683	0.122619	0.108657	0.105633	0.105633
D3	0.121848	0.121981	0.132414	0.132295	0.132295
D4	0.035997	0.035452	0.031299	0.031746	0.031746
D5	-0.054369	-0.054369	-0.059160	-0.053855	-0.053855
D6	0.105271	0.107013	0.117559	0.116093	0.116093
D7	0.080638	0.083314	0.095956	0.086899	0.086899
D8	0.014475	0.013235	0.005623	0.005204	0.005204
D9	0.050944	0.052115	0.046455	0.045915	0.045915
D10	0.091646	0.090015	0.081800	0.076840	0.076840

C. Ligands

Table 2: Descriptor values of 4 ligands which were obtained from Molfeat Server. There were many descriptors which were given as output from Molfeat Server out of which 10 descriptors were chosen for the analysis**.**

	MK0518	S1360	5CITEP	GS9137
D1	5.4266	1.8047	4.3245	8.8889
D2	0.1684	0.1589	0.1629	0.1466
D3	24.5916	12.3166	10.5133	22.2427
D4	-0.3648	-0.3461	-0.3979	-0.4087
D5	0.3959	0.3483	0.3712	0.3709
D6	0.7607	-2.7810	-2.2808	-3.9594
D7	-4.4616	2.7809	2.2808	3.9596
D8	4.4617	0.6944	0.7691	0.7796
D9	219.2070	190.3321	226.8546	143.1991
D10	0.3797	3.2499	0.8971	4.6984

PCA (Principal Components Analysis) Results Of HIV1 Integrase Sequences

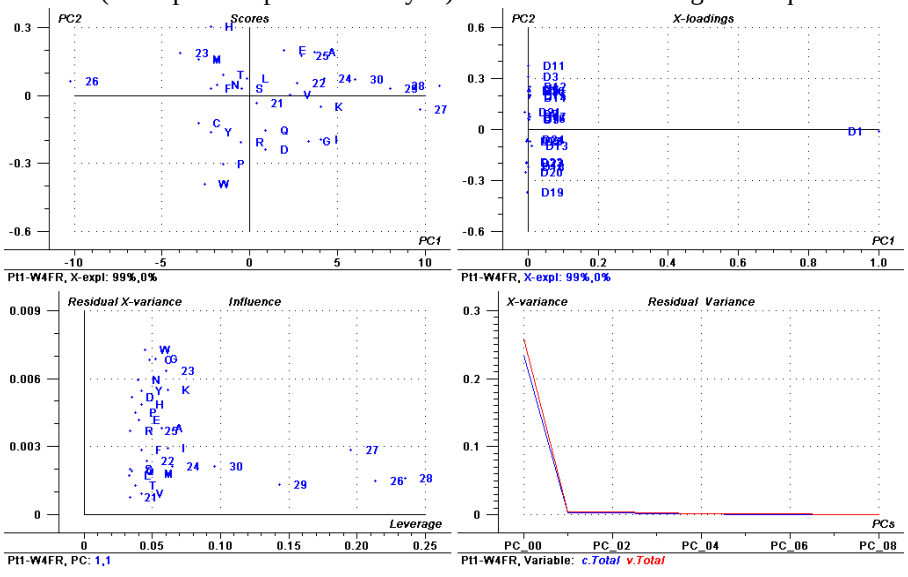


Fig.1: PCA results of sequence 1

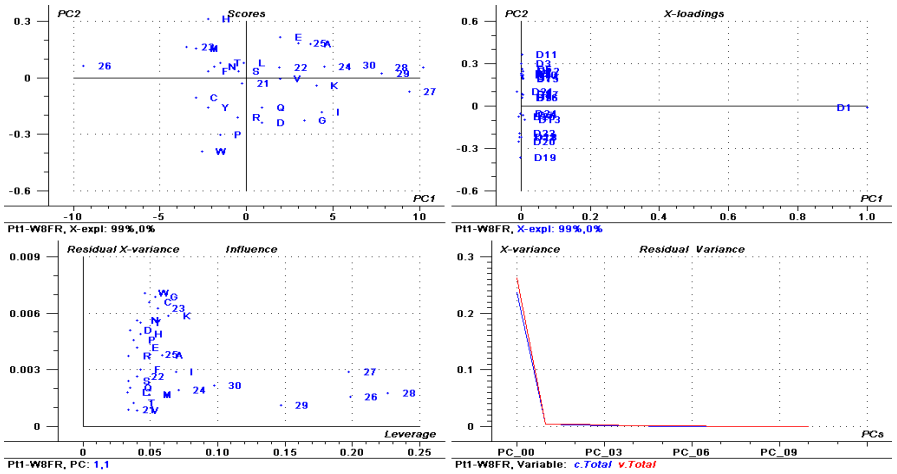


Fig.2: PCA results of sequence 2

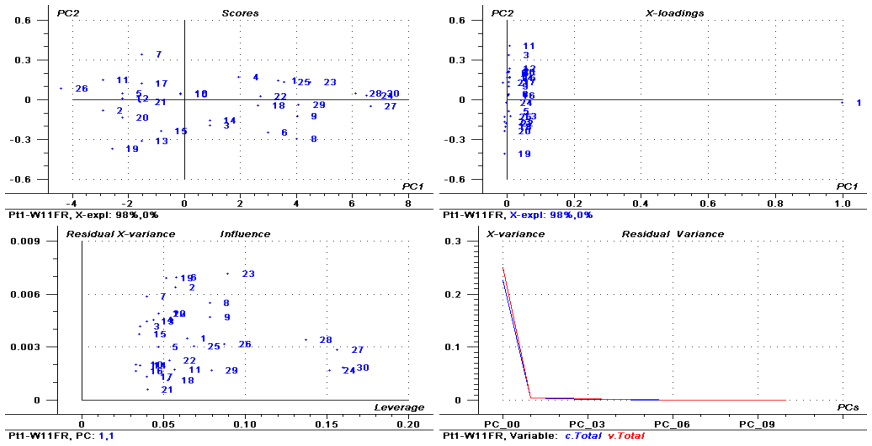


Fig.3: PCA results of sequence 3

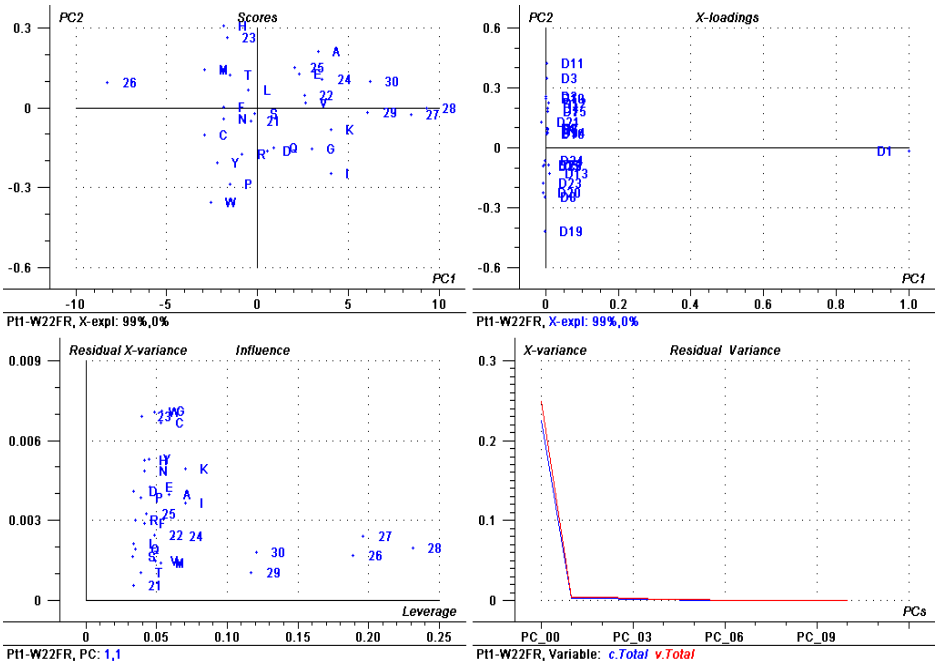


Fig.4: PCA results of sequence 4

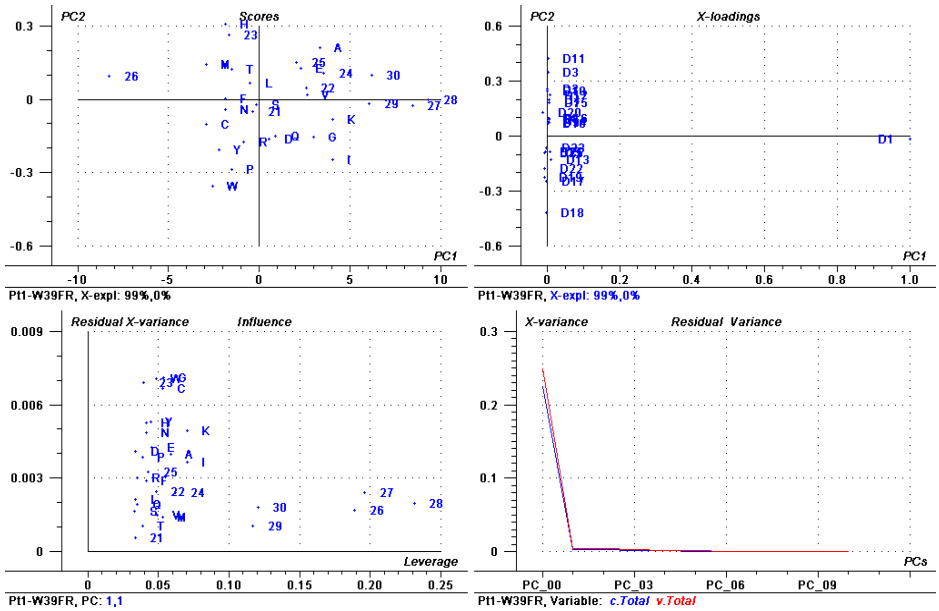


Fig.5: PCA results of sequence 5

After the PCA (Principal Component Analysis) of all the 5 sequences, 3 Principal Components from each sequence was chosen and their value was noted for the further analysis. The values of PC's are given in the table below:

Table 3: Principal Components (Z-scale Descriptors) of the 5 sequences of the HIV1 Integrase

	Principal Component 1	Principal Component 2	Principal Component 3
Seq 1	0.37345	0.30816	0.24892
Seq 2	0.36338	0.29924	0.26047
Seq 3	0.406778	0.33784	0.23486
Seq 4	0.42041	0.34829	0.25754
Seq 5	0.42041	0.34829	0.25754

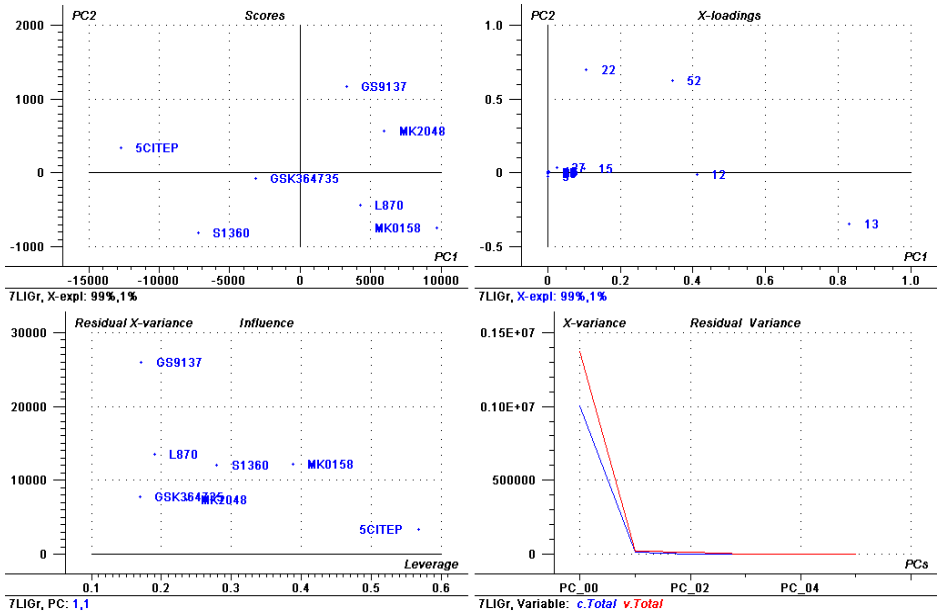


Fig.6: PCA results for ligands

Table 4: Principal Components of the ligands

Principal Component 1	Principal Component 2	Principal Component 3
0.697819	0.624323	0.032420

As the Principal Components of the sequences were found out in the same way PC's of the ligands were also found out using software Unscrambler.

The above principal components of the HIV1 Integrase were subjected to MLR (Multiple Linear Regression) and R^2 was found out using software Statgraphics. The R^2 was found out for PC's of the receptors against activities i.e. IC50 of the ligands which were found out from the literature review.

The activities of the ligands are as follows:

Table 5: Activities of the ligands

Ligands	Activity
MK0518	34nm
5CITEP	44.2nm
S1360	20nm
GS9137	8.8nm

The R^2 for PC's of the receptors against activities i.e. IC50 of the ligands is given below:

Dependent variable: ACTIVITY

Independent variables: 1, 2, 3 (PC's of the Sequences)

Parameter	Estimate	Standard Error	T-Value Statistic	P-Value
CONSTANT	640.553	132.069	4.85012	0.0400
1	-565.446	215.839	-2.61976	0.1200
2	108.649	291.366	0.372894	0.7450
3	-1647.32	343.731	-4.79249	0.0409

D. Analysis of Variance

Source	Sum of Squares	D_f	Mean Square	F-Ratio	P-Value
Model	1697.17	3	565.722	8.07	0.1123
Residual	140.283	2	70.1416		
Total (Corr.)	1837.45	5			

R-squared = 92.3653 percent = 0.9236

R-squared (adjusted for d.f.) = 80.9133 percent = 0.8091

Standard Error of Est. = 8.37506

Mean absolute error = 3.90144

Durbin-Watson statistic = 1.24109

Lag 1 residual autocorrelation = 0.179438

The R^2 for PCs of the receptors and PCs of ligands against activities i.e. IC50 of the ligands is given below:

E. Results for Multiple Regression - ACTIVITY

Dependent variable: ACTIVITY

Independent variables: 1, 2, 3 (PCs of sequences along with PCs of Ligands)

Parameter	Estimate	Standard Error	T-Value Statistic	P-Value
CONSTANT	655.89	115.246	5.69121	0.0295
1	-534.285	192.376	-2.77729	0.1089
2	57.4985	259.548	0.221533	0.8452
3	-1679.23	298.235	-5.63054	0.0301

F. Analysis of Variance

Source	Sum of Squares	D_f	Mean Square	F-Ratio	P-Value
Model	1822.06	3	607.355	11.03	0.0843
Residual	110.144	2	55.0719		
Total (Corr.)	1932.21	5			

R-squared = 94.2996 percent = 0.9429
R-squared (adjusted for d.f.) = 85.749 percent = 0.8574
Standard Error of Est. = 7.42104
Mean absolute error = 3.39239
Durbin-Watson statistic = 1.26734
Lag 1 residual autocorrelation = 0.178568

Table 6: R Squared Values

R^2 of PCs of Receptors against Ligand Activity	R^2 of PC of Receptor and Ligand (Crossterms) against Ligand Activity
0.8091	0.8574

The target was HIV1 Integrase because there is only one drug developed for the inhibition of this enzyme, which is Raltegravir (MK0158). The other drugs that were developed are in different stages of clinical trials. On the other hand, there are a number of drugs available for the inhibition of other enzymes like Protease and Reverse Transcriptase. Proteochemometrics approach was used to study the specificity and interactions of multiple HIV-1 Integrase variants with one integrase inhibitors (administered as a drug) and six other molecules which are in different stages of clinical trials, yielding a model with very good predictability and interpretability. We thoroughly analyzed the external predictive ability and the statistical significance of the model estimates, and this model can be reliably applied to the prediction and interpretation of the mechanisms of drug resistance.

Table 1 shows the descriptor values of all the 10 sequences which comprised of 50 values i.e. 10 descriptor values for 5 sequences (5 X 10 = 50). As the number of values was very high, the PCA (Principal Component Analysis) was done. The PCA was done using Unscrambler 9.7 software (Refer to Fig.1 to Fig.5 for results). Table 3 shows the Principal Components (PC) of all the 5 sequences. It comprises of 3 PCs for each sequence i.e. total there will be 15 PCs for 5 sequences (5 X 3 = 15). Now, these PCs were subjected to MLR (Multiple Linear Regression) against the activities of the ligands (Table 6), which gave us the final result i.e. R^2. R^2 calculated was between PCs of the sequences and the activities of the ligands (where the PC's were the independent variables and the activity was dependent variable) and R^2 was found to be 0.8091.

The descriptors of the ligands were also subjected to PCA and out of all the descriptor values only 3 values were chosen (Refer Table 4). MLR was done again, but this time the PCs of receptors and ligands were combined, and regression was done against the activities of the ligands. Regression analysis was done for the second time.

Here the PCs of the receptors and the ligands were independent variables and the activity of the ligands was dependent variable. The value of the R^2 was found to be **0.8574**. R^2 values indicate the accuracy of the approach to yield a robust model for development of new drugs; the higher the values are, the better the approach is.

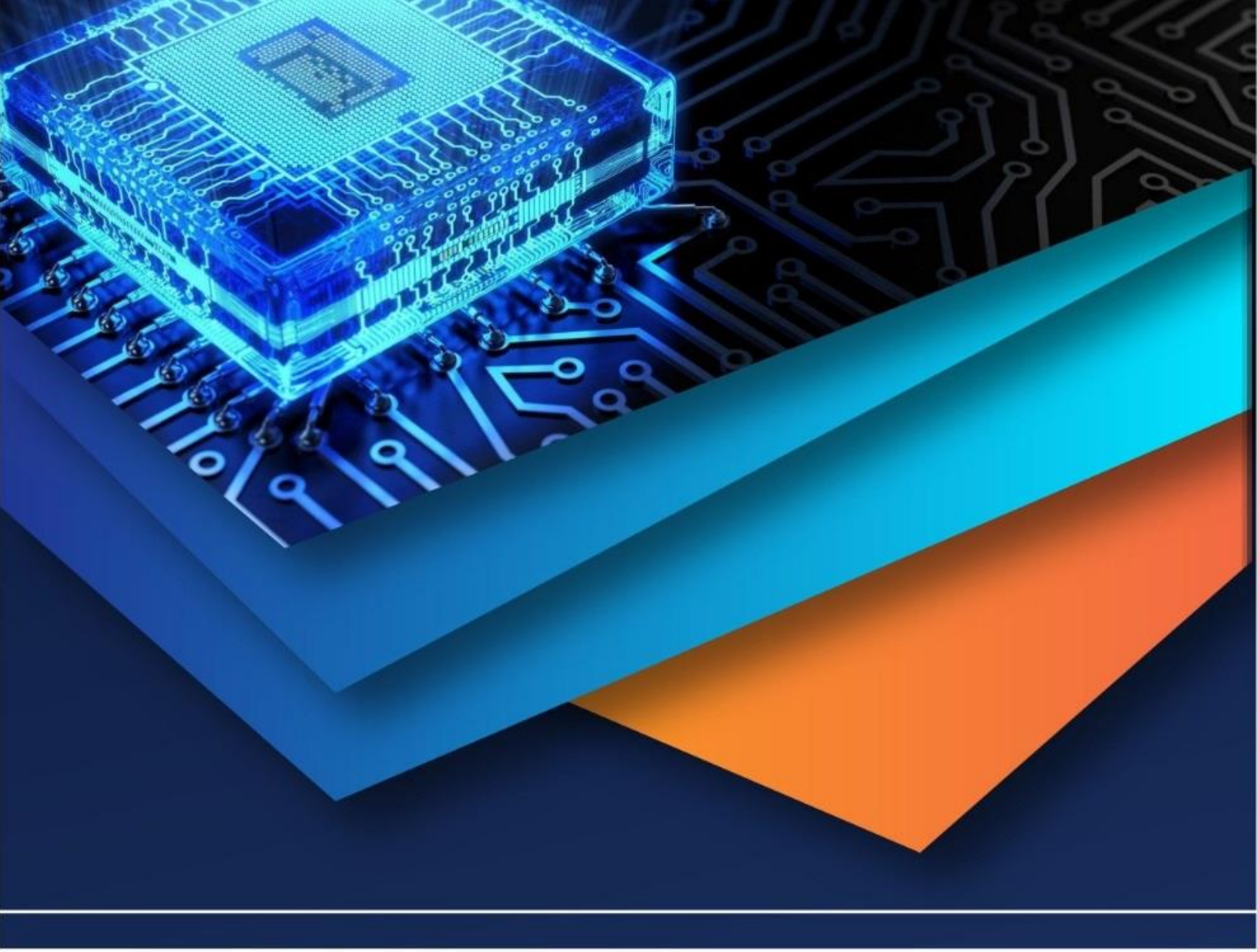
V. CONCLUSION

Proteochemometric approach compares the activity of HIV1 Integrase inhibitors including the existing drug Raltegravir, taking into consideration both, receptor as well as ligand, instead of *only* the receptor. Proteochemometric approach is hence more accurate and holds more water in order to yield a robust model for drug development, as seen by the difference between the R squared values.

This dissertation can be taken further by creating and validating an actual model by considering any of the allele of HIV1 Integrase inhibitor. In fact, the existing model can be improved upon by modifying it with the help of proteochemometrics. Also, this approach can be used to create models for important and potential targets that do not have many ligands, like cancer.

REFERENCES

[1] Robert Craigie. "HIV Integrase, a Brief Overview from Chemistry to Therapeutics" The Journal of Biological Chemistry, May 9, 2001, doi: 10.1074/jbc.R100027200.
[2] Jarl E. S. Wikberg, Martin Eklund, Ola Spjuth, Helena Strömbergsson, Torgeir Hvidsten and Maris Lapins. "Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interaction", Biochimica et Biophysica Acta (BBA) - General Subjects, Volume 1525, Issues 1–2, 16 February 2001, pp 180-190.
[3] Ivan Dimitrov, Panayot Garnev, Darren R. Flower, Irini Doytchinova. "Peptide binding to the HLA-DRB1 supertype: A proteochemometrics analysis", European Journal of Medicinal Chemistry, Volume 45, Issue 1, January 2010, Pages 236-243.
[4] Eva Freyhult, Peteris Prusis, Maris Lapinsh, Jarl ES Wikberg, Vincent Moulton and Mats G Gustafsson. "Unbiased descriptor and parameter selection confirms the potential of proteochemometric modeling", BMC Bioinformatics, 2005, Volume 6, Number 1, Page 1.
[5] Diane V. Havlir. "HIV Integrase Inhibitors — Out of the Pipeline and into the Clinic", N Engl J Med 2008; 359:416-418 DOI: 10.1056/NEJMe0804289.
[6] Raltegravir, Elvitegravir and Metoogravir. "The birth of "me-too" HIV-1 integrase inhibitors" Retrovirology, 2009, Volume 6, Number 1, Page 1.
[7] Reference Manual for PROFEAT
[8] Edward Garvey, Brian Johns, Scott Foster, Wayne Miller, Robert Ferris, Richard Hazen, Mark Underwood, Eric Boros, James Thompson, Jason Weatherhead, Cecilia Koble, Scott Allen, Lee Schaller, Ron Sherrill, Tomokazu Yoshinaga, Takeshi Noshi, Masanori Kobayashi, Akihiko Sato, Shionogi; "GSK364735 is a Potent Inhibitor of HIV Integrase and Viral Replication"; ICAAC, 47th Interscience Conference on Antimicrobial Agents and Chemotherapy; Sept 17-20, 2007 Chicago, ILL.
[9] Drugs of the Future: (http://journals.prous.com/journals/servlet/xmlxsl/pk_journals.xml_about_pr?p_JournalID=2)
[10] Anstett K, Brenner B, Mesplede T, Wainberg MA. "HIV drug resistance against strand transfer integrase inhibitors"; Retrovirology. 2017 Jun 5;14(1):36. doi: 10.1186/s12977-017-0360-7. Review.
[11] Raveendra Dayam, Tino Sanchez, Omoshile Clement, Robert Shoemaker, Shizuko Sei, and Nouri Neamati. "β-Diketo Acid Pharmacophore Hypothesis. 1. Discovery of a Novel Class of HIV-1 Integrase Inhibitors"; J. Med. Chem., 2005, 48 (1), pp 111–120.
[12] Gregg S. Jones, Fang Yu, Ameneh Zeynalzadegan, Joseph Hesselgesser, Xiaowu Chen, James Chen, Haolun Jin, Choung U. Kim, Matthew Wright, Romas Geleziunas, and Manuel Tsiang, "Preclinical Evaluation of GS-9160, a Novel Inhibitor of Human Immunodeficiency Virus Type 1 Integrase", Antimicrobial Agents and Chemotherapy [22 Dec 2008, 53(3):1194-1203]
[13] Z. R. Li, H. H. Lin, L. Y. Han, L. Jiang, X. Chen and Y. Z. Chen, "PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence"; Nucleic Acids Res. 2006 Jul 1; 34(Web Server issue): W32–W37.
[14] Reference Manual for MOLFEAT



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY