CoMFA and CoMSIA Studies on Inhibitors of HIV-1 Integrase - Bicyclic Pyrimidinones

V. RADHIKA, S. SREE KANTH and M.VIJJULATHA *

Department of Chemistry, Nizam College
Osmania University, Hyderabad- 500 001, India
drmv12002@yahoo.com

Received 9 February 2010; Accepted 21 April 2010

Abstract: To understand the structural requirements of HIV-1 integrase inhibitors and to design new ligands against human HIV-1 integrase with enhanced inhibitory potency, a 3D QSAR (quantitative structure-activity relationship) study with comparative molecular field analysis (CoMFA) and comparative molecular similarity indices analysis (CoMSIA) for a dataset of 35 bicyclic pyrimidinones which are inhibitors of human HIV-1 integrase was performed. QSAR models were computed with Sybyl. The 3D QSAR model showed very good statistical result, namely q^2 , r^2 and r^2_{pred} values were high for both CoMFA and CoMSIA. Based on the high values for q^2 and r^2 we are confident that the 3D QSAR model gives good predictions that may be used to design better HIV-1 integrase inhibitors. The CoMFA and CoMSIA models reveal that steric and electrostatic fields contribute significantly with biological activities of the studied compounds.

Keywords: QSAR, CoMFA, CoMSIA, Bicyclic pyrimidinones.

Introduction

Human immunodeficiency virus type 1 (HIV-1) is responsible for human acquired immunodeficiency syndrome (AIDS), one of the most urgent world health threats. The HIV-1 genome, together with protease and reverse transcriptase encodes a third enzyme, integrase, which is necessary for the viral life cycle and is an appealing target for drug discovery for its essential direct counterpart in the host cells^{1,2}. HIV-1 integrase catalyses the incorporation of the viral DNA into the host cell's genome through a multi-step process. In the initial assembly step, integrase binds to specific sequences of the double stranded viral DNA, newly formed by reverse transcription. An endonucleolytic cleavage of the last two nucleotides from the 3' terminus of both strands of the viral DNA generates a processed DNA/integrase pre-integration complex, which crosses the nuclear membrane to access the host genome. The cell DNA is then cleaved by integrase in a nonspecific way and the shortened strands of viral DNA are integrated into the host DNA sequence.

Selective inhibition of the HIV-1 integrase activity causes an interruption of the HIV-1 replication cycle and could represent an answer to the unmet medical need for new and improved treatments³.

It is reported that, bicyclic pyrimidinones⁴⁻⁶ have been evolved from *N*-methyl pyrimidinone. Introduction of a suitably substituted amino moiety modulated the physical and chemical properties of the molecules and conferred nanomolar activity in the inhibition of spread of HIV-1 infection in cell culture.

Quantitative structure-activity relationships (QSAR) models establish statistical relationships between the biological activity exerted by a series of compounds and a set of parameters determined from their structures (structural descriptors). The mathematical 3D QSAR equations can be computed with the help of a large number of statistical models such as multi-linear regression, partial least squares (PLS) or artificial neural networks. An important component of all QSAR models is a proper validation and evaluation of the prediction power.

CoMFA and CoMSIA are usually applied in combination with experimental techniques when molecules are highly flexible^{7,8}. In the present study molecules are not as flexible and theoretical calculations are sufficient for their alignment.

The CoMFA and CoMSIA methodologies assume that a suitable sampling of steric, electrostatic and hydrogen-bond donor fields around a set of aligned molecules provides all the information necessary for understanding their biological properties. The present study is aimed to gain insights into the steric, electrostatic and hydrogen-bond donor properties of bicyclo pyrimidinones, their influence on the activity and to derive predictive 3D QSAR models for design and prediction of the activities of new derivatives for this class of inhibitors.

Experimental

The *in vitro* biological activity data reported as IC₅₀ for inhibition of human HIV-1 integrase by the bicyclic pyrimidinones was taken from the published work of Muraglia *et al.*⁹ and used for the current study. The reported IC₅₀ values were converted into the corresponding pIC₅₀ values. In the present study there are four core molecules with different substituents comprising of 35 bicyclic pyrimidinones. Their structures, pIC₅₀ and the predicted activities from CoMFA and CoMSIA were given in the Table 1(a) & (b).

Three dimensional structure building and all modeling were performed using the Sybyl 6.9 molecular modeling program package¹⁰. Energy minimization was performed using the Tripos force field¹¹ and the Gasteiger-Huckel charge with a distance dependant dielectric and conjugate gradient algorithm with convergence criterion of 0.001 k cal/mol.

Scheme 1. Structure of the compound BP1-BP35

Table 1(a). Structures and biological activities of molecules used in the present study

Compound	R ¹	R ²	pIC ₅₀ (Observed)	CoMFA		CoMSIA	
				pIC ₅₀ (Predicted)	Residual	pIC ₅₀ (Predicted)	Residual
BP1	-	H	7.222	7.228	0.006	7.354	0.132
BP2	-		7.000	6.972	-0.028	6.926	-0.074
BP3	CH ₃		7.222	7.261	0.039	7.129	-0.093

Table 1 (b). Structures and biological activities of molecules used in the present study

Compound	n	R	pIC ₅₀ (Observed)	CoMFA		CoMSIA	
				pIC ₅₀ (Predicted)	Residual	pIC ₅₀ (Predicted)	Residual
BP4	0		8.1549	8.201	0.0461	8.127	-0.0279
BP5	1		7.1611	7.190	0.0289	7.251	0.0899
BP6	1		7.2006	7.190	-0.0106	7.264	0.0634
BP7	1		8.3979	8.414	0.0161	8.381	-0.0169
BP8	2		7.2076	7.199	-0.0086	7.181	-0.0266
BP9	2		7.2076	7.188	-0.0196	7.114	-0.0936
*BP10	0	H	8.0969	8.080	-0.0169	8.173	0.0761
BP11	1	H	8.3010	8.167	-0.134	8.130	-0.171
BP12	2	H	7.7212	7.830	0.1088	7.979	0.2578
BP13	1		7.9208	7.781	-0.1398	7.847	-0.0738
*BP14	1		8.1549	7.921	-0.2339	8.287	0.1321

Contd...

*BP15	1		8.3010	8.392	0.091	8.440	0.139
BP16	2		8.000	7.976	-0.024	8.024	0.024
BP17	1		7.8538	7.816	-0.0378	7.914	0.0602
*BP18 ^a	2		8.2218	7.650	-0.5718	7.744	-0.4778
BP19	1		8.1549	8.116	-0.0389	8.246	0.0911
BP20	1		8.0457	8.113	0.0673	8.072	0.0263
BP21	1		7.9586	7.922	-0.0366	8.061	0.1024
BP22	1		8.1544	8.121	-0.0334	8.141	-0.0134
BP23	0		7.699	7.741	0.042	7.749	0.05
BP24	1		8.1549	8.130	-0.0249	7.996	-0.1589
BP25	2		7.886	7.877	-0.009	7.848	-0.038

Compd	n	R ¹	R ²	pIC ₅₀ (observed)	CoMFA		CoMSIA	
					pIC ₅₀ (predicted)	Residual	pIC ₅₀ (predicted)	Residual
BP26	0	--CH ₃	--CH ₃	8.057	8.099	0.042	8.024	-0.033
BP27	1	--CH ₃	--CH ₃	8.2218	8.208	-0.0138	8.083	-0.1388
*BP28	2	--CH ₃	--CH ₃	8.0969	7.995	-0.1019	8.158	0.0611
BP29	1		--CH ₃	7.7447	7.904	0.1593	7.906	0.1613
BP30	1		--CH ₃	8.5228	8.542	0.0192	8.460	-0.0628
BP31	1		--CH ₃	8.000	8.014	0.014	8.045	0.045
BP32	1		--CH ₃	8.0457	8.025	-0.0207	8.003	-0.0427
*BP33	1		--CH ₃	8.2218	8.163	-0.0588	8.064	-0.1578
*BP34	2		--CH ₃	7.7695	7.678	-0.0915	7.552	-0.2175
BP35	1		H	8.1509	8.142	-0.0089	8.114	-0.0369

Compounds with * represent the test set. 'a' stands for the outlier molecule

Alignment

In the present study the optimized structures were aligned on the template BP30, which is the most active molecule among the set (Figure 1). All the molecules were aligned by the Align Database command available in Sybyl using maximum substructure. This adjusts the geometry of the molecules such that their steric and electrostatic fields match the fields of the template molecule.

CoMFA analysis

The steric and electrostatic CoMFA¹² potential fields were calculated at each lattice intersection of a regularly spaced grid of 2.0Å. The grid box dimensions were determined automatically in such a way that region boundaries were extended beyond 4.0Å in each direction from coordinates of each molecule.

The Van der Waals potentials and columbic terms were calculated using the Tripos force field. A sp³ hybridized carbon atom with +1 charge served as probe atom to calculate steric and electrostatic fields. The steric and electrostatic contributions were truncated to +30.0 k cal/mol.

Figure 1. Atom based alignment of bicyclic pyrimidinones on BP30

CoMSIA analysis

In the CoMSIA¹³ model, three physicochemical properties, namely steric, electrostatic and hydrogen-bond donor fields were calculated using a sp^3 hybridized carbon probe atom with a +1 charge and a radius of 1.0Å placed at regular grid spacing of 2.0Å and a similar lattice box as used in CoMFA calculations.

Partial least square (PLS) analysis

The CoMFA or CoMSIA descriptors were used as independent variables and pIC_{50} values as dependant variables in partial least square regression analysis¹⁴⁻¹⁷. The minimum sigma (column filtering) was set to 2.0 k cal/mol. to improve the signal-to-noise ratio. The predictive correlation coefficient (r^2_{pred}) based on the test molecules, is computed with the formula $r^2_{pred} = (SD-PRESS) / SD$, where SD is the sum of the squared deviations between the biological activities of the test set and mean activities of training set molecules and PRESS is the sum of squared deviation between predicted and actual activity for every molecule in test set. r^2_{cv} is also calculated, where $r^2_{cv} = (SD'-PRESS) / SD'$, where SD' is the sum of the squared deviations between the biological activities of the test set and mean activities of test set molecules.

Results and Discussion

Validation of the 3D QSAR models

To validate the stability and predictive ability of the obtained models, 7 compounds not included in the construction of CoMFA and CoMSIA models were selected as test set. Thus the data base of 35 molecules consisted of 28 molecules in the training set and 7 in the test set. The plots of experimental activity *vs.* predicted activity (Figure 2) clearly show that the predicted pIC_{50} values of the test set compounds are in good agreement with the experimental data in a tolerable error range, with r^2_{pred} of 0.549 and 0.546 for CoMFA and CoMSIA models, respectively. These results derived from CoMFA and CoMSIA models show stable models that can reliably used to design novel inhibitors with desired activity for the HIV-1 integrase.

Significance of statistical parameters

The statistical parameters are given in Table 2, r^2 value explains a variance in biological activity with 98.1% and 94.7% (CoMFA and CoMSIA respectively) for training set and r^2_{pred} value explains a variance in biological activity with 54.9% and 54.6% (CoMFA and CoMSIA respectively) for test set. The steric and electrostatic parameters contribute 48.2% and 51.8% towards biological activity by CoMFA. Steric, electrostatic and hydrogen-bond donor parameters contribute 34.0%, 50.5% and 15.5% respectively towards biological activity by CoMSIA.

Figure 2. Plots of experimental activity vs. predicted activity from CoMFA and CoMSIA (Pink colored values represent the test set)

Table 2. The statistical parameters for the CoMFA and CoMSIA models

Statistical Parameters	CoMFA Model	CoMSIA Model
Molecules in training set	28	28
Molecules in test set	7	7
q^{2a}	0.561	0.699
r^{2b}	0.981	0.947
ONC ^c	6	5
SEE ^d	0.067	0.110
F^{e}_{ratio}	177.051	79.147
r^{2f}_{pred}	0.549	0.546
r^{2g}_{cv}	0.257	0.426
Fraction of Field Contributions:		
Steric	0.482	0.340
Electrostatic	0.518	0.505
Hydrogen-bond donor	-----	0.155

^aCross validated correlation coefficient, ^bNon-cross validated correlation coefficient, ^cOptimum number of components, ^dStandard error of estimate, ^eF-test value, ^fPredictive r^2 , ^gPredictive r^2 by taking mean of test set activities

CoMFA and CoMSIA contour maps

To visualize the information content of the derived 3D QSAR models, CoMFA and CoMSIA contour maps were generated¹⁸. The field energies at each lattice point were calculated as the scalar results of the coefficient and the standard deviation associated with a particular column of the data table ("stdev*coeff"), which was always plotted as the percentage of the contribution to the CoMFA or CoMSIA equation.

The CoMFA contour maps of steric and electrostatic field are shown in Figures 3 and 4. The contours of the steric map are shown in yellow and green and those of the electrostatic map are shown in red and blue. Greater values of 'Bio-Activity Measurement' are correlated with: more bulk near green; less bulk near yellow; more positive charge near blue and more negative charge near red. Steric contour maps of BP30 (high pIC_{50}), BP17 (medium pIC_{50}) and BP2 (least pIC_{50}) are displayed in Figure 3. Less activity of molecule BP2 can be explained by presence of bulky groups like N-methyl groups into the disfavored region and the moderate activity of BP17 is due to the penetration of bulky groups from unfavored to favored region. BP30 higher activity can be explained because it is orienting away from the disfavored region and is inclining towards the favored region.

Electrostatic contour maps of BP30, BP17 and BP2 are displayed in Figure 4. In BP17, electronegative groups like nitrogen, oxygen are orienting towards both favoring and unfavouring regions which explains its moderate activity. In molecule BP2, the electronegative groups are orienting towards unfavouring region, so it shows least activity. In BP30, the electronegative groups are orienting away from unfavouring or blue region. Blue contour also has hydrogens which are on the electronegative groups. Both the conditions are favorable. Hence, BP30 shows highest activity.

Contour maps of CoMFA help to design molecules with increased activity. Introduction of bulkier groups or hydrophobic groups on the pyrazole ring in BP30, suggests increased activity.

Figure 3. Steric contour maps of BP30 (a), BP17 (b), BP2 (c) molecules from CoMFA

Figure 4. Electrostatic contour maps of BP30 (a), BP17 (b), BP2 (c) molecules from CoMFA

The steric and electrostatic field distributions of CoMSIA are depicted in Figure 5 and 6 and are generally in accordance with the field distributions of CoMFA map (Figure 3 & 4)

Figure 5. Steric contour maps of BP30 (a), BP17 (b), BP2 (c) molecules from CoMSIA

Figure 6. Electrostatic contour maps of BP30 (a), BP17 (b), BP2 (c) molecules from CoMSIA

The CoMSIA contour map of the hydrogen-bond donor is shown in Figure 7. Grey contours (contribution level of 80%) indicate regions where hydrogen-bond donor group increases activity. In the molecule BP30 (a), the donor contours are overlapping the donor hydrogens (hydrogens on nitrogen and oxygen which can be substituted easily). Whereas in the molecule BP2 (b), the donor contours do not overlap the donor hydrogens as they are away from the field. This also explains the high activity of BP30 and least activity of BP2.

Figure 7. Hydrogen-bond donor contour maps of BP30 (a), BP2 (b) molecules from CoMSIA

HIV-1 integrase (PDB ID-1qs4) is not a complete protein. Some amino acids are missing. The structure has to be generated and loop modeling has to be done. Once a valid protein structure is obtained inhibitors will be docked into the active site. Depending upon their pharmacodynamic characteristics new molecules can be designed that will be synthesized to analyze their activity against HIV-1 integrase.

Conclusion

In this study, we have established predictive CoMFA and CoMSIA 3D QSAR models for the HIV-1 integrase inhibitors. Compounds like BP30, which has better activity can be taken as best hit molecule based on which new inhibitors of HIV-1 integrase can be designed and used for further studies like docking analysis, synthetic studies and biological activity studies.

Acknowledgments

We greatly thank Dr. G. Narhari Shastry, Scientist, Indian Institute of Chemical Technology (IICT) for allowing us to use Sybyl software. We are thankful to Department of Chemistry, Nizam College for providing us all the help required. UGC is greatly thanked for their financial support.

References

1. Young S D, *Curr Opin Drug Discovery Dev.*, 2001, **4**, 402.
2. (a) Esposito D, Craigie R, *Adv Virus Res.*, 1999, **52**, 319-333; (b) Asante-Appiah E and Skalka A.M, *Adv Virus Res.*, 1999, **52**, 351-369.
3. Pommier Y, Johnson A.A and Marchand C, *Nat Rev Drug Discovery*, 2005, **4**(3), 236-248.
4. Summa V, Petrocchi A, Matassa V.G, Gardelli C, Muraglia E, Rowley M, Paz O.G, Laufer R, Monteagudo E and Pace P, *J Med Chem.*, 2006, **49**(23), 6646-6649.
5. Petrocchi A, Koch U, Matassa V.G, Pacini B, Stillmock K.A and Summa V, *Bioorg Med Chem Lett.*, 2007, **17**(2), 350-53.
6. Pace P, Di Francesco M.E, Gardelli C, Harper S, Muraglia E, Nizi E, Orvieto F, Petrocchi A, Poma M, Rowley M, Scarpelli R, Laufer R, Gonzalez P.O, Monteagudo E, Bonelli F, Hazuda D, Stillmock K.A and Summa V, *J Med Chem.*, 2007, **50**, 2225.

7. Serdar Durdagi, Manthos G. Papadopoulos, Demetris P. Papahatjis and Thomas Mavromoustakos, *Bioorg Med Chem Lett.*, 2007, **17**(24), 6754-6763.
8. Serdar Durdagi, Manthos G. Papadopoulos, Demetris P. Papahatjis and Thomas Mavromoustakos, *J. Med Chem.*, 2007, **50**(12), 2875-2885.
9. Ester Muraglia, Olaf Kinzel, Cristina Gardelli, Benedetta Crescenzi, Monica Donghi, Marco Ferrara, Emanuela Nizi, Federica Orvieto, Giovanna Pescatore, Ralph Laufer, Odalys Gonzalez-Paz, Annalise Di Marco, Fabrizio Fiore, Edith Monteagudo, Massimiliano Fonsi, Peter J. Felock, Michael Rowley and Vincenzo Summa, *J Med. Chem.*, 2008, **51**, 861.
10. SYBYL version 6.9, Tripos Associates, St. Louis (MO) 1999.
11. Clark M, Cramer R D III and Van Opdenbosch N, *J Comput Chem.*, 1989, **10**, 982-1012.
12. Cramer R.D III, Patterson D.E and Bunce J.D, *J Am Chem Soc.*, 1988, **110**, 5959-5967.
13. Klebe G, Abraham U and Mietzner T, *J Med Chem.*, 1994, **37**, 4130-4146.
14. Cramer R.D III, Bunce J.D, Patterson D.E and Frank I.E, *Quant Struct Act Relat.*, 1988, **7**, 18-25.
15. Geladi P, Kowalski B.R, *Anal Chem Acta*, 1986, **185**, 1.
16. Baroni M, Clementi S, Cruciani G, Costantino G, Riganelli D and Oberrauch E, *J Chemometrics*, 1992, **6**, 347-356.
17. Wold S, Ruhe A, Wold H and Dunn W.J.I, *SIAM J Sci Stat Comput.*, 1984, **5**, 735.
18. Juan Du, Jin Qin, Huanxiang Liu and Xiaojun Yao, *J Mol Graph Model*, 2008, **27**(2), 95-104.