International Journal of Pure and Applied Researches http://ijopaar.com; 2017 Vol. 2(2); pp. 111-120, ISSN: 2455-474X

QSAR Study and Determination of More Potent Peptidic HIV-1-Protease Inhibitors

Anil Kumar Soni
Department of Chemistry, Shia P.G. College, Lucknow, U.P., INDIA
Email: anilsony82@gmail.com

Abstract

QSAR study of three sets of peptidic HIV-protease inhibitors has been studied. The descriptor used for study is log P. For QSAR modeling, log P has been calculated using the atom-typing scheme of Ghose and Crippen associated with CAChe. From the structure activity relationship discussion it is clear that log P is an important parameter for QSAR study. Compounds-16, 18, 19, 20, 21, 22, 24, 25, 27, 29 and 30 have high inhibitory activity greater than 9.2 but their log P value is more than 5.0 hence these compounds do not follow minimal hydrophobicity principle and show poor pharmacokinetics. While, compounds-07, 09, 11, 13, 23 and 28 have inhibitory activity ranging from 8.11 to 9.51 and their log P value is less than 5.0, hence these compounds follow minimal hydrophobicity principle and show good pharmacokinetics. Finally QSAR modeling has been made using log P as a descriptor. The QSAR model $A_{Predicted}$ = 0.620278 log P + 5.53898 has reliable predictive power as it has rCV^2 = 0.566415 and r^2 = 0.59635. On the basis of this QSAR model one can propose theoretical formalism of new peptidic HIV-protease inhibitors that show better pharmacokinetics, including higher oral bioavailability and slow excretion.

Keywords: QSAR, Protease inhibitor, Log P, Pharmacokinetics

1. Introduction

Anti-HIV drug discovery has been increasingly focusing on HIV-protease as a potential therapeutic target. This enzyme is required for cleavage and virion maturation (steps of viral life cycle) to change the non-infectious progeny to infectious ones (*Kohl et al., 1988*), (*Peng et al., 1989*). Thus, the prevention of this HIV-protease enzyme plays a crucial role in the development of anti-HIV chemotherapy (*Jacabson et al., 1995*). Since HIV-protease is an aspartic protease and its substrate is peptide in nature, hence a number of peptide derived compounds have been identified as HIV-protease inhibitors (*Romines and Thaisrviongs, 1995*). Based on log P, the QSAR of 41 protease inhibitors is presented in this paper. The QSAR is based on the premise that there is a relatively simple mathematical relationship between the biological activity of a drug and its physico-chemical properties (*Hansch and Fujita, 1964*), (*Hou et al., 1999*). For instance if the hydrophobicity of a drug is important for its biological activity, then changing the substituents on the drug so as to alter its hydrophobicity will affect its activity (*Hansch et al., 1987*). Of course, the biological activities of these drugs (peptidic HIV-protease inhibitors) depend on their hydrophobicities. A measure of the drug`s hydrophobicity is its partition coefficient (P) between two immiscible solvents, octanol and water at equilibrium (*Leviton and Barken, 1972*). Biological activity may be expressed as 1/C, where C is the drug concentration required to achieve a specified level of biological function and can be therefore expressed as Log (1/C) = k_1(log P) + k_2. Here $k_1$and k_2 are constants, where optimum values in this QSAR can be determined by computerized curve-fitting methods. For compounds with a larger range of log P values it is better described by quadratic equation Log (1/C) = k_1 (log P)^2 + k_2 log P + k_3 (*Leo et al., 1971*).

2. Materials and Methods

The study materials of this paper are presented in three set. The first set comprises of derivatives of urea isostere and the second and third sets derivatives of other isosteres (*Getman et al., 1993*), (*Holloway et al., 1995*). For QSAR prediction, the 3D modeling and geometry optimization of

all the derivatives have been done with the help of CAChe pro software by opting PM3 methods (*Stewart, 1989*). Multiple regression analysis has been made by using Project Leader Software associated with CAChe (*Soni et al., 2011*).

3. Results and Discussion

On the basis of skeleton structure of parent compound the peptidic HIV-protease inhibitors are divided in three sets (Fig. 1-3), which along with their biological activity is presented in Tables 1-3. Log P has been calculated using the atom-typing scheme of Ghose and Crippen associated with CAChe and has been presented in Tables 4-6 (*Ghose and Crippen, 1986*). Each table has been divided into subgroups in order to demonstrate better and sequential relationship between the biological activity and reactivity parameters. The QSAR study of each set has been discussed as below.

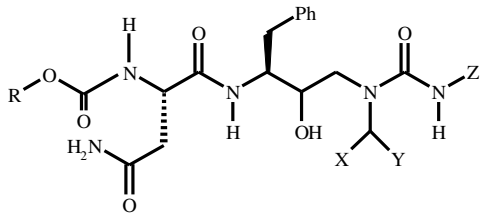


Fig.1. *Skeleton structure of parent compound of first set*

First Set

The first set consists of fifteen urea isostere derivatives and their biological activity has been measured in terms of inhibitory activity (*Getman et al., 1993*). The reactivity indices along with biological activity of this set of compounds are placed in Table 4. A close look at this table indicates that addition of alkyl /hydrophobic group increases the activity. The examination of Table 4 also indicates that there is direct relationship between log P and inhibitory activity. Although there is a direct relationship but there is no sequential rise or fall. In order to provide sequential relationship we divided the set into five subgroups- A, B, C, D and E. In subgroup-A, compounds-01, 04 and 05; in subgroup-B, compounds-14, 06 and 08; in subgroup-C, compounds-03 and 10; in subgroup-D, compounds-15, 07 and 09; and in subgroup-E, compounds-02 and 11 show the sequential relationship very clearly. Compounds-12 and 13 do not follow the sequential trend.

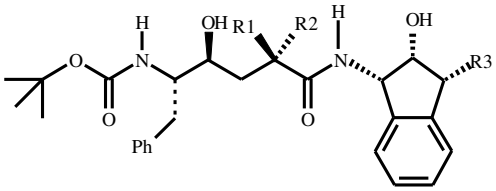


Fig. 2. *Skeleton structure of parent compound of second set*

Second Set

Second Set of derivatives also contains fifteen compounds and their biological activity is also shown in term of inhibitory activity (*Holloway et al., 1995*). The biological activity

and the reactivity indices of these derivatives are given in Table 5. A close look at this table indicates that addition of alkyl, aryl and other hydrophobic group increases activity. The discussion also indicates that there is a direct relationship between log P and inhibitory activity of this set of compounds. We have divided the compounds of this table into three different subgroups-A, B and C. Subgroup-A and B contain four compounds but the subgroup-C contains two compounds. If the relationship is examined in these three subgroups separately the sequence is also exhibited clearly. In subgroup-A, compounds-26, 17, 16 and 20; in subgroup-B, compounds-23, 22, 19 and 30; and in subgroup-C, compounds-28 and 29 show the direct relationship very clearly. Compounds-27, 24 and 25 on the other hand follow inverse relationship.

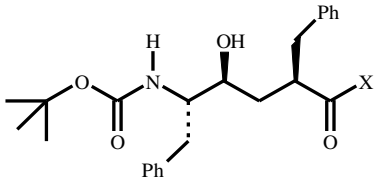


Fig. 3. *Skeleton structure of parent compound of third set*

Third Set

Third set of derivative contain eleven compounds and their biological activity is also shown in term of inhibitory activity (*Holloway et al., 1995*). The activities alongwith reactivity indices are given in Table 5. Examination of this table shows that the biological activity is inversely proportional to log P. It is further observed that the addition of hydrophilic substituents increases activity. The inverse relationship can be better represented if the compounds of this set are divided in to three subgroups-A, B and C. Subgroup-A includes the compounds-33, 36 and 34; subgroup-B, includes the compounds-40, 39 and 37; and the subgroup-C includes the compounds-35, 38 and 41. Compounds-31 and 32 do not follow the sequential trend.

Pharmacokinetics

Since, the pharmacokinetics of a drug is as important to its efficacy as is its pharmacodynamics, both must be optimized in producing a medicinally useful drug (*Olson et al., 1993*). The one of the most important empirically based rule formulated by Christopher Lipinski is that a compound is likely to exhibit poor absorption or permeation if its value of log P is greater than 5 (*Lipinski, 2004*). Thus, the most effective drugs are usually a compromise; they are neither too hydrophobic nor too hydrophilic. Compounds-16, 18, 19, 20, 21, 22, 24, 25, 27, 29 and 30 show poor pharmacokinitics, including low oral bioavailability and rapid excretion but these have high inhibitory activity which is greater than 9.2 (*Olson et al., 1993*).

According to Huff, the inhibitor-enzyme binding is dominated by hydrophobic interaction (*Huff, 1991*). Protease inhibitors bind to the protease binding site pocket that has a considerable number of hydrophobic residues (*Fitzgerald et al., 1990*), (*Erickson et al., 1990*). The residues that make up these pockets are Valine-32, Isolucine-47, Isolucine-50 and Isolucine-84 in each monomer. The amino acids Isolucine and Valine have 4.5 and 4.2 hydropathy index and are the top hydrophobic amino acids respectively. The receptor cleft or pocket may not be completely homogeneous (hydrophobic) that is, there must be some hydrophilic residues because the compounds-07, 09, 11, 13, 23 and 28 although have log P value (2.07 to 4.60) lower than 5.0 yet they have higher inhibitory activity from 8.11 to 9.51. It is further proved by inverse relationship between log P and activity as shown by compounds of third set.

QSAR Modeling

Regression analysis is made with log P and regression equation $A_{Predicted}$ = 0.620278 log P + 5.53898 having rCV^2 = 0.566415 and r^2 = 0.59635 been obtained. The predicted inhibitory activity of various derivatives as obtained from above regression equation has been presented in Table-7. A reference to these tables clearly indicates that predicted activities are close to observed activity as residual value (Δ) is small.

4. Conclusions

From the above structure activity relationship discussion it is clear that log P is an important parameter for QSAR study. Compounds-16, 18, 19, 20, 21, 22, 24, 25, 27, 29 and 30 have high inhibitory activity greater than 9.2 but their log P value is more than 5.0 hence these compounds do not follow minimal hydrophobicity principle and show poor pharmacokinetics. While, compounds-07, 09, 11, 13, 23 and 28 have inhibitory activity ranging from 8.11 to 9.51 and their log P value is less than 5.0, hence these compounds follow minimal hydrophobicity principle and show good pharmacokinetics. Finally QSAR modeling has been made using log P as a descriptor. The QSAR model $A_{Predicted}$ = 0.620278 log P + 5.53898 has reliable predictive power as it has rCV^2 = 0.566415 and r^2 = 0.59635. On the basis of this QSAR model one can propose theoretical formalism of new peptidic HIV-protease inhibitors that show better pharmacokinetics, including higher oral bioavailability and slow excretion.

5. Acknowledgement

The author is thankful to Principal, M.L.K. (P.G.) College, Balrampur-271201 (U.P.), for laboratory facilities.

Referances

1. Erickson, J., Neidhart, D. J., Van Drive, J., Kempf, D. J., Wang, X. C., Norbeck, D. W., Plattner, J. J., Rittenhouse, J. W., Turon, M., Wideburg, N., Kohlbrenner, W. E., Simmer, R., Helfrich, R., Paul, D. A. and Knigge, M. 1990; *Science* 249, 527.
2. Fitzgerald, P. M. D., McKeever, B,M., Van Middlesworth, J. F., Springer, J. P., Heimbach, J. C., Leu, C. T., Herber, W. K., Dixon, R. A. F. and Darke, P. L. 1990; *J. Biol. Chem.* 265, 14209.
3. Getman, D. P., Decrescenzo, G. A., Heintz, R. M., Reed, K. L., Talley, J. J., Bryant, M. L., Clare, M., Houseman, K. A., Marr, J. J., Mueller, R. A., Vazquez, M. L., Shieh, H. S. Stallings, W. C.; and Stegeman, R. A. 1993, *J. Med. Chem.* 36, 288.
4. Ghose, A. K. and Crippen, G. M.1986; *J. Comput. Chem.* 7, 565.
5. Hansch, C. and Fujita, T. 1964; *J. Am. Chem. Soc.* 86, 1616.
6. Hansch, C., Bjorkroth, J. P. and Leo, A. 1987; *J. Pharm. Sci.* 76, 663.
7. Hou, T. J., Wang, J. M., Liao, N. and Xu, X. J. 1999; *J. Chem. Inf. Comput. Sci.* 39, 775.
8. Holloway, M. K., Wai, J. M., Halgren, T. A., Fitzgerald, P. M. D., Vacca, J. P., Dorsey, B. D., Levin, R. B., Thompson, W. J., Chen, L. J., deSolms, S. J., Gaffin, N., Ghosh, A. K., Giuliani, E. A., Graham, S. L., Guare, J. P., Hungate, R. W., Lyle, T. A., Sanders, W. M., Tucker, T. J., Wiggins, M., Wiscoust, C. M., Woltersdof, O. W., Young, S. D., Darke, P. L. and Zugay, J. A. 1995; *J. Med. Chem.* 38, 305.
9. Huff, J. R. 1991; *J. Med. Chem.* 34, 23.
10. Jacabson, H., Yasargil, K., Winslow, D. L., Craig, J. C., Krohn, A., Duncan, I. B. and Mous, J. 1995; *J. Virol.* 206, 327.

11. Kohl, N. E., Emini, N. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon R. A. F., Scolnick, E. M. and Sigal, I.S. 1988; *Proc. Natl. Acad. Sci. U.S.A.* 85, 4686.
12. Leo, A., Hansch, C. and Elkins, D. 1971; *Chem. Rev.* 71, 525.
13. Leviton, H. and Barken, J. L. 1972; *Science (Washington, D.C.)* 176, 1423.
14. Lipinski, C. A. 2004; *Drug Discov. Today: Technologies 1*, 337.
15. Olson, G. L., Bolin, D. R., Bonner, M. P., Bos, M., Cook, C. M., Fry, D. C., Graves, B. J., Hatada, M., Hill, D. F., Khan, M., Madison, V. S., Rusiedci, V. K., Sarabu, R., Sepinwall, J.' Vincet, G. P. and Voss, M. F. 1993; *J. Med. Chem.* 36, 3039.
16. Peng, C., Ho, B.K., Chang, T.W. and Chang, N.T. 1989; *J. Virol.* 63, 2550.
17. Romines, K. R. and Thaisrviongs, S. 1995; *Drugs Future* 20, 377.
18. Soni, A.K., Sahu, V. K., Singh, P. P. 2011; *Asian J. Chem.* 23, 671.
19. Stewart, J. J. P. 1989; *J. Comp. Chem.* 10, 209.

Table-1. First set of derivatives containing 15 compounds and their biological activity in terms of inhibitory activity

Compound	*Substituents*				*Activity*
No.	*R*	*X*	*Y*	*Z*	
1	Cbz	H	$CHMe_2$	Me	5.82
2	Qua	H	$CHMe_2$	n-Bu	6.90
3	Cbz	H	$CHMe_2$	n-Pr	6.29
4	Cbz	H	$CHMe_2$	Et	6.48
5	Cbz	H	$CHMe_2$	i-Pr	6.59
6	Cbz	H	$CHMe_2$	t-Bu	7.46
7	Qua	H	$CHMe_2$	t-Bu	8.22
8	Cbz	H	CH_2CHMe_2	t-Bu	7.89
9	Qua	H	CH_2CHMe_2	t-Bu	8.52
10	Cbz	H	C_6H_{11}	t-Bu	7.54
11	Qua	H	C_6H_{11}	t-Bu	8.30
12	Cbz	H	C_6H_5	t-Bu	7.72
13	Qua	H	C_6H_5	t-Bu	8.52
14	Cbz	H	4-Py	t-Bu	6.98
15	Qua	H	4-Py	t-Bu	7.72

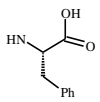
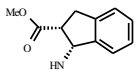
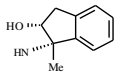
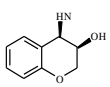
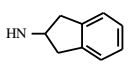
Cbz = Carbobenzyloxy, Qua = Quinolinyl-2-Carboxamide

Table-2. Second set of derivatives containing 15 compounds and their biological activity in terms of inhibitory activity

Compound	*Substituents*			*Activity*
No.	R_1	*R2*	*R3*	
16	CH_2Ph	H	H	9.60
17	CH_2Ph	Me	H	8.11
18	CH_2CH_2Ph	H	OH	9.72

19	CH_2-4-CF_3Ph	H	H	9.59
20	(E)CH_2CH=CHPh	H	H	9.64
21	$CH_2C_6F_5$	H	H	9.22
22	CH_2-4-CH_3Ph	H	H	9.54
23	CH_2-4-NH_2Ph	H	H	9.51
24	CH_2-4-NO_2Ph	H	H	9.57
25	CH_2-4-OHPh	H	H	9.80
26	CH_2CH=CH_2	H	H	7.56
27	CH_2-4-IPh	H	H	9.14
28	CH_2C(O)Ph	H	H	8.27
29	CH_2SPh	H	H	9.60
30	CH_2-4-CMe_3Ph	H	H	9.77

Table-3. Third set of derivatives containing 11 compounds and their biological activity in terms of inhibitory activity

Compound No.	Substituents (X)	Activity
31	Ph CH_2NH-	6.94
32		7.47
33		6.16
34		6.79
35		7.18
36		6.67
37		6.91
38		7.39
39		6.89

40		6.84
41		7.41

Table-4. Calculation of log P and its relationship with activity of first set of derivatives containing 15 compounds

Compound No.	*Log P*	*Activity*
Subgroup A		
1	1.558	5.82
4	1.901	6.48
5	2.314	6.59
Subgroup B		
14	1.643	6.98
6	2.392	7.46
8	2.716	7.89
Subgroup C		
3	2.369	6.29
10	3.72	7.54
Subgroup D		
15	1.664	7.72
7	2.549	8.22
9	3.077	8.52
Subgroup E		
2	2.922	6.9
11	3.359	8.3
12*	2.908	7.72
13*	2.07	8.52

Table-5. Calculation of log P and its relationship with activity of second set of derivatives containing 15 compounds

Compound No.	*Log P*	*Activity*
Subgroup A		
26	4.352	7.56

16	5.388	9.6
17	4.994	8.11
20	6.317	9.64
Subgroup B		
23	4.605	9.51
22	5.855	9.54
19	6.271	9.59
30	7.015	9.77
Subgroup C		
28	4.461	8.27
29	5.175	9.6
Subgroup D		
27	6.646	9.14
24	5.436	9.57
25	5.104	9.8
18*	5.784	9.72
21*	6.086	9.22

** indicates the compound do not follow the trend*

Table-6. Calculation of log P and its relationship with activity of third set of derivatives containing 11 compounds

Compound No.	*Log P*	*Activity*
Subgroup A		
33	3.489	6.16
36	3.367	6.67
34	2.784	6.79
Subgroup B		
40	4.145	6.84
39	3.913	6.89
37	2.649	6.91
Subgroup C		
35	4.08	7.18
38	3.732	7.39
41	3.359	7.41
31*	3.609	6.94

32*	4.058	7.47

** indicates the compound do not follow the trend*

Table-7. Predicted activity of all the 41 compounds as obtained from regression equation, $A_{Predicted}$

Compound No.	*Activity*	*Predicted Activity*	*Δ*
1	5.82	6.51	0.69
2	6.90	7.35	0.45
3	6.29	7.01	0.72
4	6.48	6.72	0.24
5	6.59	6.97	0.38
6	7.46	7.02	0.44
7	8.22	7.12	1.10
8	7.89	7.22	0.67
9	8.52	7.45	1.07
10	7.54	7.85	0.31
11	8.30	7.62	0.68
12	7.72	7.34	0.38
13	8.52	6.82	1.70
14	6.98	6.56	0.42
15	7.72	6.57	1.15
16	9.60	8.88	0.72
17	8.11	8.64	0.53
18	9.72	9.13	0.59
19	9.59	9.43	0.16
20	9.64	9.46	0.18
21	9.22	9.31	0.09
22	9.54	9.17	0.37
23	9.51	8.40	1.12
24	9.57	8.91	0.66
25	9.80	8.71	1.10
26	7.56	8.24	0.68
27	9.14	9.66	0.52
28	8.27	8.31	0.04
29	9.60	8.75	0.85

30	9.77	9.89	0.12
31	6.94	7.78	0.84
32	7.47	8.06	0.59
33	6.16	7.70	1.54
34	6.79	7.27	0.48
35	7.18	8.07	0.89
36	6.67	7.63	0.96
37	6.91	7.18	0.27
38	7.39	7.85	0.46
39	6.89	7.97	1.08
40	6.84	8.11	1.27
41	7.41	7.62	0.21