Pergamon

Bioorganic & Medicinal Chemistry Letters 12 (2002) 3453–3457

BIOORGANIC & MEDICINAL CHEMISTRY LETTERS

Six-Membered Cyclic Ureas as HIV-1 Protease Inhibitors: A QSAR Study Based on CODESSA PRO Approach

Alan R. Katritzky,[a,*] Alexander Oliferenko,[a] Andre Lomaka[b] and Mati Karelson[b]

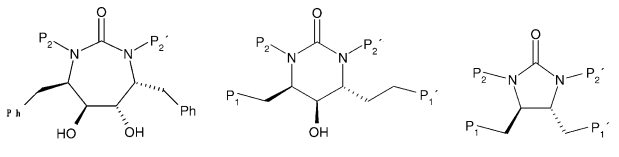
[a]*Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, PO Box 11720, Gainesville, FL 32611-7200, USA*

[b]*Department of Chemistry, University of Tartu, 2 Jakobi Str., Tartu, Estonia*

Received 25 April 2002; accepted 14 August 2002

Abstract—Quantitative structure–activity relationships (QSAR) for HIV-1 protease inhibitory activity of substituted tetrahydropyrimidinones have been produced using CODESSA PRO methodology and software. The best four-parameter equation ($R^2_{cv} = 0.847$) allowed us to reveal two main structural factors which are strongly correlated with the title activity: molecular hydrophobicity and ability to form hydrogen bonds with the target enzyme.

Efficient inhibition of aspartyl protease can decrease HIV-1 virulence.[1] This enzyme is therefore a major target for structure based inhibitor design.[2,3] The first inhibitors designed to fit the substrate-binding groove of HIV-1 proteases were peptide-derived species. However, such compounds possess poor pharmacokinetics[3] and are complex and expensive to synthesize. To overcome these difficulties, significant efforts have been devoted to the development and SAR study of non-peptide protease inhibitors such as cyclic ureas and other heterocycles.[4–6] Examples of such inhibitors include C_2-symmetric hexahydro-1,3-diazepin-2-ones (**1**), which are complementary to the C_2-symmetric protease of HIV, tetrahydropyrimidin-2-ones (**2**), and imidazolidin-2-ones (**3**) (Fig. 1). Compounds of types **2** and **3** can be prepared from the seven-membered diazepinones **1** by ring contraction.[5] In a SAR study of tetrahydropyrimidinones **2**,[4] it is stated that the phenethyl P1′ substituent, the hydroxyl group, and the urea carbonyl are crucial for high inhibitory activity. By varying the P1′ substituent as well as P2 and P2′, the authors sought out and found some patterns of the target activity.



1 2 3

Figure 1. Structures of substituted seven-, six-, and five-membered cyclic ureas, potent HIV-1 protease inhibitors.

Because search by synthesis and testing is time consuming and expensive, several groups have attempted to develop theoretical models for the analysis and prediction of inhibitory potency. To classify existing peptidomimetic aspartyl protease inhibitors, Luciana et al.[7] performed a multivariate statistical analysis of a set of 24 target compounds in the hyperspace of eight theoretical descriptors such as heat of formation, O–H and N–H bond lengths, dipole moment, charge at a selected carbonyl oxygen atom, HOMO energy, molecular radius, and calculated log P. Use of such statistical techniques as the PCA (principal component analysis) based SIMCA (soft independent modeling of class analogy) method and the *k*-th nearest neighbors (kNN) clustering allowed them to partition all 24 structures, with few errors, into three tentative classes: good, moderate, and poor inhibitors.

A linear response model has been developed for the prediction of log (IC_{50}) values of diaminodiol-based peptidomimetics using molecular mechanical complexation

*Corresponding author. Tel.: +1-352-392-0554; fax: +1-352-392-9199; e-mail: katritzky@chem.ufl.edu

0960-894X/02/$ - see front matter © 2002 Elsevier Science Ltd. All rights reserved.
PII: S0960-894X(02)00741-2

energy, side chain surface area, and a rigidity parameter as descriptors.[8] The recent work of Gupta et al.[9] on tetrahydropyrimidinones **2** proposes various QSAR models. One topological index (Kier and Hall first order valence connectivity) and two indicator parameters (flags for presence or absence of certain groups) were used as independent variables. Separate three-parameter models were suggested for compounds with similar or dissimilar tetrahydropyrimidinone ring *N*-substituents: the model for 26 symmetrically substituted inhibitors had $R^2=0.694$; $s=0.640$; $F=17$, while that for 25 dissimilar compounds had $R^2=0.821$; $s=0.570$; $F=32$. The authors concluded that the inhibitory activity of the title compounds depends primarily on the hydrophobic properties of P2 and P2′ substituents expressed in terms of the first order valence connectivity index. While the size and shape of a molecule can be adequately expressed in terms of topological indices, consideration of molecular bulk properties alone is insufficient for a proper modeling of protease inhibition, unless the charge distribution and hydrogen bonding are treated explicitly.

The current work aimed to use CODESSA methodology to derive QSAR models of the protease inhibitory activity of tetrahydropyrimidinones, accounting for all significant structural information and applicable to all cyclic ureas. Using a large pool of theoretically derived descriptors and multiple regression analysis as a statistical tool, and a combined set of 51 compounds, we present in the following sections a short outline of CODESSA methodology and optimum one-, two-, three-, and four-parameter equations which explain the inhibitory activity of tetrahydropyrimidinones **2** towards aspartyl proteases.

The training set for the present investigation was built on tetrahydropyrimidinone inhibition activities obtained experimentally by De Lucca et al.[4] and further subjected to QSAR analysis by Gupta et al.[9] In distinction to their work,[9] we treated symmetrical and nonsymmetrical compounds as a single data set. Molecules were modeled using the MM+ method of HyperChem.[10] Final optimizations were performed with MOPAC[11] using the AM1[12] semiempirical method. About 600 constitutional, topological, geometrical, charge-related, semiempirical, and molecular-, atomic-, bond-type descriptors were calculated with a CODESSA PRO software package.[13] The applicability of CODESSA methodology to various QSAR/QSPR problems has been convincingly demonstrated in a series of our recent publications.[14,15]

In accordance with the previous treatments of HIV-1 protease inhibitory activity,[4,6] two main structural factors correlate with the target activity: (i) molecular size and shape, and (ii) electronic distribution including the ability for hydrogen bonding. Hydrogen bond descriptors such as the fractional positive surface area of hydrogen donor species ($^{H}FPSA^{(2)}$), and some modifications of the latter, occur in all the QSAR models produced with rather large positive coefficients and Student *t*-values, indicating the eminent importance of hydrogen bond formation during ligand–receptor binding. This is perfectly in line with the data extracted by the previous authors[4] from X-ray diffraction experiments who found that the hydroxyl group of **2** (Fig. 1) is involved in hydrogen bonding interactions with the catalytic aspartic acid residues Asp 25/125.

The importance of molecular size and shape in QSAR studies of tetrahydropyrimidinones has previously been demonstrated clearly.[9] In our treatment of tetrahydropyrimidinone protease inhibitory activity we have used a topological descriptor based on Shannon's expression of information content[16] as a term representing molecular size and shape. Structural information content of order 1 (^{1}SIC) accounts for both molecular constitutional and structural diversity. Despite being the only descriptor in eq 1, ^{1}SIC shows quite a significant correlation with $R^2=0.646$. Figure 2 displays the corresponding correlation plot of observed versus predicted values of the inhibitory activity of substituted tetrahydropyrimidinone.

$$\log(1/K_i) = 2.139(\pm 0.740) + 0.167(\pm 0.018)\,{}^{1}SIC$$
$$R^2 = 0.646;\ R^2_{cv} = 0.617;\ s = 0.750;\ F = 90. \quad (1)$$

This measure of structural complexity along with the hydrogen bond descriptor, $^{HD}FPSA^{(2)}$ (fractional positive surface area of hydrogen donors),[17] produced a two-parameter model (eq 2) for all the 51 tetrahydropyrimidinones, superior to those previously obtained[9] for two separate sets of 26 symmetrical ($R^2=0.694$; $s=0.640$; $F=17$) and 25 nonsymmetrical compounds ($R^2=0.821$; $s=0.570$; $F=32$):

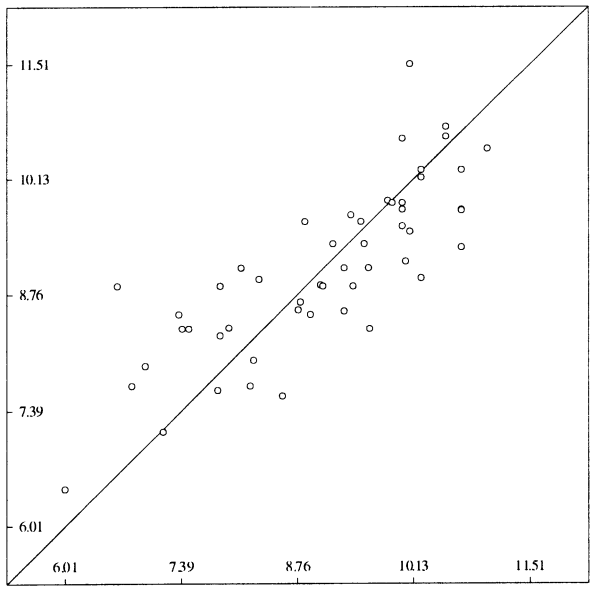


Figure 2. Correlation plot of observed versus predicted values of the inhibitory activity of substituted tetrahydropyrimidinones. One-parameter eq 1.

$$\log(1/K_i) = 2.168(\pm 0.535) + 0.108(\pm 0.010)\,{}^{2}SIC + 12.750(\pm 1.848)\,{}^{HD}FPSA^{(2)}$$
$$R^2 = 0.816;\ R^2_{cv} = 0.792;\ s = 0.545;\ F = 107 \qquad (2)$$

Our new software, CODESSA PRO, enables the search and selection of correlations in a very fast, automated regime.[18] Adding more parameters significantly improves the correlation. In particular, the shape of a molecule, an extremely important feature during ligand-receptor binding, can be satisfactorily expressed in terms of molecular principal axes. In the case of the substituted tetrahydropyrimidinones under study, the first moment of inertia takes its highest values for the nonsymmetrical species, providing an apparent discrimination between the symmetrical and nonsymmetrical ones. The best three-parameter eq 3 includes the hydrogen bond descriptor ${}^{HD}FCPSA^{(2)}$ which is a charge weighted analogue of ${}^{HD}FPSA^{(2)}$ (weighted by Zefirov's atomic charges[19]), the first moment of inertia (I^A), and the Balaban topological index J that is a measure of molecular 'centricity'. The latter descriptor can be considered as complementary to I^A, since both of them describe in different ways the nonuniformity of mass distribution in a molecule. Both I^A and the Balaban index bear the negative sign, pointing to the superiority of C_2-symmetric structures with identical P2 and P2′ substituents, which is opposite to the conclusion of higher efficacy of asymmetric inhibitors reported earlier.[9]

$$\log(1/K_i) = 15.901(\pm 0.917) + 261.881(\pm 24.823)\,{}^{HD}FCPSA^{(2)} - 160.948(\pm 36.184)I^A - 5.708(\pm 0.746)J$$
$$R^2 = 0.855;\ R^2_{cv} = 0.832;\ s = 0.489;\ F = 92 \qquad (3)$$

The four-parameter model (see Fig. 3) by eq 4, includes an important new descriptor, the minimum electron-nuclear attraction energy for O–H bond, E^{min}_{e-n}(O−H), which may be regarded as an immediate scale of the bond strength. The negative contribution made by this descriptor is an additional argument in favor of the high importance of the hydrogen bonding within the active sites of aspartyl proteases. Increased acidic dissociation of O–H bond (oxime or alcoholic, see Fig. 1 and Table 1) probably facilitates the formation of the binding complex with the receptor. As can be seen from Table 1, two data points, compounds **3** and **17**, have deviations higher than one logarithmic unit. We could suppose that the latter has the intramolecular hydrogen bond between the amino group and fluorine substituent of the benzylic P_2 (P_2') group, which cannot be accounted by the hydrogen bond descriptor ${}^{HD}FPSA^{(2)}$, whereas for the former outlier, compound 3, the moment of inertia term I^A is underestimated and, hence, the log $(1/K_i)$ is overestimated (this can be supported by a partial correlation with I^A, which is not given here).

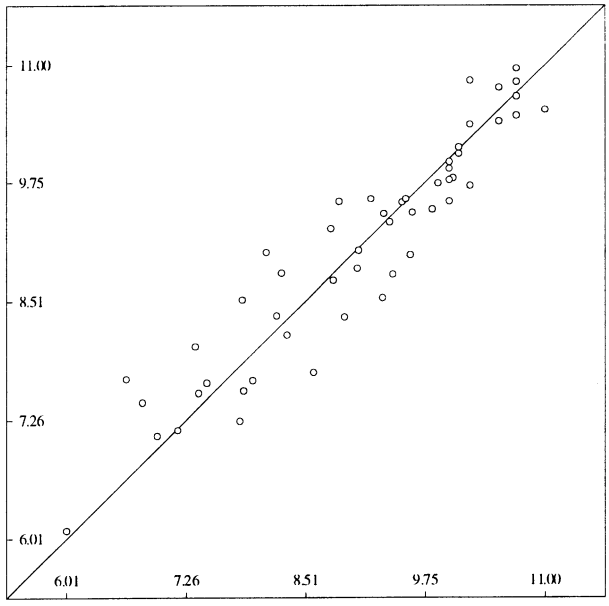


Figure 3. Correlation plot of observed versus predicted values of the inhibitory activity of substituted tetrahydropyrimidinones. Four-parameter eq 4.

$$\log(1/K_i) = 42.585(\pm 8.445) + 15.601(\pm 1.776)\,{}^{HD}FPSA^{(2)} - 141.775(\pm 34.923)I^A - 6.204(\pm 0.732)J - 0.234(\pm 0.073)E^{min}_{e-n}(O-H)$$
$$R^2 = 0.873;\ R^2_{cv} = 0.847;\ s = 0.463;\ F = 79 \qquad (4)$$

Theoretical CODESSA PRO QSAR analysis of six-membered cyclic ureas with respect to their HIV-1 protease inhibitory potency demonstrates that complex biological activity can be adequately described in terms of theoretical descriptors if a sufficient selection of descriptors is available. For the purpose of the current study, CODESSA PRO software enabled us to construct hundreds of correlations and select automatically the best ones. Three important structural features responsible for the tetrahydropyrimidinones' protease inhibitory activity are revealed. First, inhibitor molecules should have rather bulky and yet sufficiently polar symmetrical P2 and P2′ substituents; molecules with nonsymmetrical P2 and P2′ substituents, especially with a missing P2′ group are inferior inhibitors. Second, inhibitory species should possess hydrogen donor groups that facilitate and strengthen binding with proteases. Of minor importance is the electronic complementarity that can be tracked as ability for electrostatic interactions inside the receptor socket.

All the QSAR models were subjected to validation using the 'leave-one-out' technique, the cross-validation results are estimated by cross-validation squared correlation coefficient, R^2_{cv}. Because of the small size of the dataset, we did not divide it into a training set and a test set.

Table 1. P2/P2′ substitution of tetrahydropyrimidinones and their observed and predicted with eq 4 values of aspartyl protease inhibitory activity

ID	Substituents	Inhibitory activity, $\log(1/K_i)$		
	P2/P2′	Exp.	Calcd	Difference[a]
1	3-Cyanobenzyl	7.959	7.656	0.303
2	3-Cyanobenzyl[b]	7.854	7.695	0.159
3	3-Cyano-4-fluorobenzyl	6.638	7.702	−1.064
4	3-Acetylbenzyl	9.040	8.674	0.366
5	3-Hydroxymethylbenzyl	9.309	8.724	0.585
6	3-Carboxybenzyl	9.060	9.242	−0.182
7	3-(Carboxamido) benzyl	10.045	9.553	0.492
8	3-(Carboxamido) benzyl[b]	9.508	9.382	0.126
9	3-(Carboxamido) benzyl[c]	9.823	9.285	0.538
10	3-(Carboxamido)-4-fluorobenzyl	8.853	9.386	−0.533
11	3-(Carboxamido oxime)benzyl	10.698	10.767	−0.069
12	3-(Carboxamido oxime)benzyl[b]	10.698	10.670	0.028
13	3-(Carboxamido oxime)benzyl[c]	11.000	10.599	0.401
14	3-(Carboxamido oxime)-4-fluoro benzyl	10.221	10.689	−0.468
15	3-Aminobenzyl	8.769	9.469	−0.700
16	3-Amino-4-fluorobenzyl	9.602	9.340	0.262
17	4-Amino-3-fluorobenzyl	8.096	9.354	−1.258
18	3-(*N*-Methylamino)-4-fluorobenzyl	8.309	8.306	0.003
19	3-(Pyrazol-3-yl) benzyl	10.000	10.488	−0.488
20	Indazol-5-yl-methyl	10.698	10.294	0.404
21	Indazol-6-yl-methyl	10.221	10.297	−0.076
22	(3-Methylindazol-5-yl)methyl	10.000	9.638	0.362
23	3-Aminoindazol-5-yl-methyl	10.522	10.137	0.385
24	3-Aminobenzisoxazol-5-yl-methyl	9.387	9.921	−0.534
25	3-(5-Methyl-2-pyridil carboxamido)-benzyl	10.096	10.430	−0.334
26	3-(*N*-2-Thiazolyl carboxamido)-benzyl	10.522	10.390	0.132
27	H/H	6.013	6.005	0.008
28	Benzyl/H	7.180	7.189	−0.009
29	3-Cyanobenzyl/ H	7.823	6.942	0.881
30	3-Cyano-4-fluorobenzyl/H	6.958	6.918	0.040
31	3-Hydroxylbenzyl/H	8.585	8.095	0.490
32	3-Aminobenzyl/H	8.207	8.510	−0.303
33	3-(Carboxamido) benzyl/ H	8.251	8.510	−0.259
34	3-(Carboxamido)-4-fluorobenzyl/H	7.853	8.297	−0.444
35	3-(Carboxamido oxime)benzyl/H	9.619	9.271	0.348
36	3-Aminoindazol-5-yl-methyl/H	9.420	8.840	0.580
37	Indazol-5-yl-methyl/H	8.920	8.511	0.409
38	H/Indazol-5-yl-methyl	7.356	8.134	−0.778
39	H/3-Aminobenzyl	6.804	7.475	−0.671
40	Benzyl/3-cyano-4-fluorobenzyl	7.481	7.468	0.013
41	3-Cyano-4-fluorobenzyl/benzyl	7.397	7.505	−0.108
42	3-(Carboxamido oxime)benzyl/cyclopropylmethyl	9.318	9.629	−0.311
43	Benzyl/3-aminoindazol-5-yl-methyl	9.187	9.424	−0.237
44	3-Aminoindazol-5-yl-methyl/benzyl	9.552	9.424	0.128
45	Indazol-5-yl-methyl/3-aminobenzyl	10.096	9.997	0.099
46	3-Aminobenzyl/indazol-5-yl-methyl	10.000	9.743	0.257
47	3-Aminobenzyl/(3-methylindazol-5-yl)methyl	9.886	9.749	0.137
48	3-Aminobenzyl/3-cyanobenzyl	8.795	8.671	0.124
49	3-Aminobenzyl/3-aminoindazol-5-yl-methyl	10.000	9.823	0.177
50	3-Aminobenzyl/3-(carboxamido) benzyl	10.221	9.707	0.514
51	3-Aminobenzyl/3-(carboxamido oxime)benzyl	10.698	10.622	0.076

[a]Difference = exp.−calcd
[b]The P1/P1′ phenyl groups bear a 4-fluoro-substituent.
[c]The P1/P1′ phenyl groups bear 3- and 4-fluoro-substituents.

References and Notes

1. Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. *Science* **1989**, *246*, 1149.
2. West, M. L.; Fairlie, D. P. *Trends Pharm. Sci.* **1995**, *16*, 67.
3. Garg, R.; Gupta, S. P.; Gao, H.; Babu, M. S.; Debnath, A. K.; Hansch, C. *Chem. Rev.* **1999**, *99*, 3525.
4. De Lucca, G. V.; Liang, J.; De Lucca, I. *J. Med. Chem.* **1999**, *42*, 135.
5. De Lucca, G. V. *Bioorg. Med. Chem. Lett.* **1997**, *7*, 495.
6. Baures, P. W. *Org. Lett.* **1999**, *1*, 249.
7. Figuerido, L. J. O.; Antunes, O. A. C. *Int. J. Quant. Chem.* **1999**, *76*, 744.
8. Tossi, A.; Bonin, I.; Antcheva, N.; Norbedo, S.; Benedetti, F.; Miertus, S.; Nair, A. C.; Maliar, T.; Bello, F. D.; Palu, G.; Romeo, D. *Eur. J. Biochem.* **2000**, *267*, 1715.
9. Gayathri, P.; Pande, V.; Sivakumar, R.; Gupta, S. P. *Bioorg. Med. Chem.* **2001**, *9*, 3059.
10. *HyperChem 5.1*; HyperCube Inc., 1999.
11. Stewart, J. J. P. *MOPAC Program Package*; QCPE 1989, No. 455.
12. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

13. http://www.codessa-pro.com
14. Maran, U.; Karelson, M.; Katritzky, A. R. *Quant. Struct-Act. Relat* **1999**, *18*, 3.
15. Katritzky, A. R.; Maran, U.; Lobanov, V. S.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1.
16. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; Wiley: New York, 2000.
17. Stanton, D. T.; Jurs, P. C. *Anal. Chem.* **1990**, *62*, 2323.
18. Katritzky, A. R.; Perumal, S.; Petrukhin, R. *J. Org. Chem.* **2001**, *66*, 4036.
19. Zefirov, N. S.; Kirpichenok, M. A.; Izmailov, F. F.; Trofimov, M. I. *Dokl. Akad. Nauk (Engl. Transl.)* **1987**, *296*, 883.