Int. J. Chem. Sci.: 11(2), 2013, 911-924
ISSN 0972-768X
www.sadgurupublications.com

ROLE OF MOLECULAR REDUNDANCY IN MODELING OF CLOSTRIDIUM HISTOLYTICUM COLLAGENASE INHIBITORY ACTIVITY OF SULFONYL-1-AMINE HYDROXAMATES

SHWETA (THAKUR) RATHORE*, M. L. GANGWAL and DINESH PALIWAL

P. M. B. Gujarati Science College, INDORE (M.P.) INDIA

ABSTRACT

The present research paper deals with the role of molecular redundancy in modeling of CLOHI collagenase inhibitors. The regression analysis has shown that out of the pool of topological indices used, the MRI is the best one for protease inhibitor properties and this has been discussed on the basis of various statistical parameters.

Key words: MRI, CLOHI, Regression analysis, Topological index.

INTRODUCTION

Clostridium histolyticum (CLOHI) is an anaerobic, spore forming, bacterium that is present in soil and feces. The ammonia and proteases it produces, including several collagenases, digest proteins outside its body into amino acids, which it eats. When *C. histolyticum* infects an open wound, it can also necrotize tissue for consumption by secreting an exotoxin that induces cytolysis. It is the primary pathogen in cases of gas gangrene. Collagenase *clostridium histolyticum*, a collagenase produced by this species, is applied for the medical treatment of chronic conditions characterized by excessive collagen deposition. It is used as a powder-and-solvent injection kit for the treatment for Dupuytren's contracture, a condition where the fingers bend towards the palm and cannot be fully straightened.

Collagenase is an enzyme that catalyzes the hydrolysis/degradation/depolymerizing collagen and gelatin. An aromatic sulfonyl-1-amine hydroxamates inhibits matrix metalloprotease activity. Matrix metalloproteinase (MMPs), an increasing family of zinc-

*Author for correspondence; E-mail: shweta_rathore02@yahoo.com

and calcium-dependent endopeptidases, are involved in both the tissue remodeling and the degradation of extracellular matrix (ECM). These enzymes have been a pharmaceutical target for over 25 years in order to develop many families of therapeutically important synthetic matrix metalloproteinase inhibitors (MMPIs) for the treatment of several serious pathologies. We herein discuss the hydroxamic acid-based MMPIs with respect to their, structure–activity relationship (SAR) and quantitative structure–activity relationship (QSAR).

Quantitative structural-activity relationship methodology can be helpful in stretching of large library of possible drug candidates for selectivity and potency. Mathematical models are formed that correlate molecular structure to an activity or property of interest. Molecular structure is encoded through the generation of the descriptions, which are numerical values corresponding to topological, geometric, or electoral structural features. Molecular Redundancy Index rank molecule according to symmetry and to include structural characteristics influencing biological activity. MRI is derived from information theory and molecular graph theory[32] and is defined as –

$$MRI = \frac{\sum n_i \log n_i}{N \log N} \quad \ldots(1)$$

Where n_i is the number of atoms of the same kind in the i^{th} atom set, i is the number of different atoms in the molecule and N is the molecular negentropy.

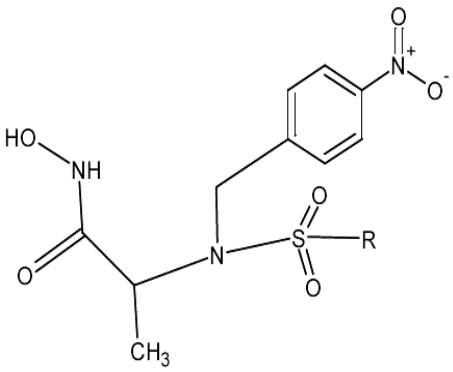
Eq. (1) shows that calculation of MRI leads to quantification of the information content. It encodes the salient steric properties of the molecules in cases where biological activity is not specific. It ranks them correctly according to nonspecific biological potency and thus, provides mechanistic interpretation of drugs at molecular level based on probability consideration.

EXPERIMENTAL

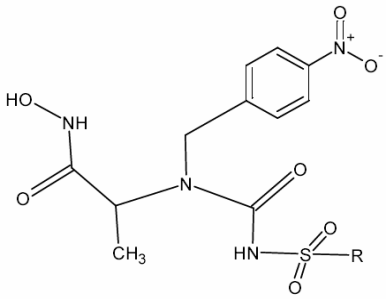
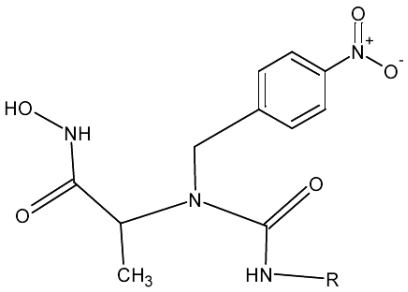
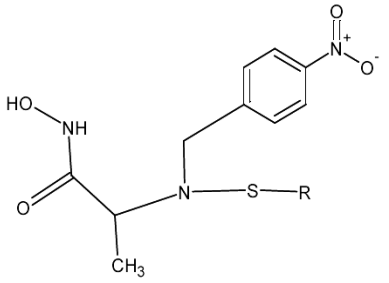
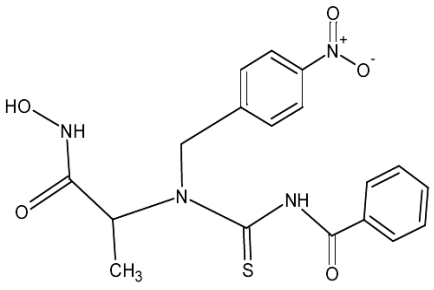
Methodology

The methodology used in the present investigation is to model the protease inhibitory activities of sulfonyl-1-amine hydroxamates using molecular redundancy as a molecular descriptor. The structural details of the parent compounds are given in Table 1, from these compounds, a large set of substituted hydroxamates derivatives will be obtained (Table 1) and the role of molecular redundancy in modeling their inhibition potential will be investigated. Multiple Linear Regression calculates QSAR equations by performing standard multivariable regression calculations using multiple variables in a single equation.

Table 1:

Comp.	Parent structure of comp.	R
1		$-CH_3$
2		$-CF_3$
3		$-CCI_3$
4		$-C_4F_9$
5		$-n-C_8F_{17}$
6		Me_2N-
7		$-C_6H_5$
8		$-CH_2C_6H_5$
9		$4-F-C_6H_4-$
10		$4-Cl-C_6H_4-$
11		$4-Br-C_6H_4-$
12		$4-I-C_6H_4-$
13		$4-CH_3-C_6H_4-$
14		$4-NO_2-C_6H_4-$
15		$3-NO_2-C_6H_4-$
16		$2-NO_2-C_6H_4-$
17		$3Cl-4\ NO_2-C_6H_3-$
18		$4Ac-NH-C_6H_4-$
19		$4Ac-C_6H_4-$
20		$3\ CF_3-C_6H_4-$
21		$2,5-\ Cl_2-C_6H_3-$
22		$4\ CH_3O-C_6H_4-$
23		$2,4,6(CH_3)_3-C_6H_2-$
24		$2\ HO-3,5\ Cl_2-C_6H_2-$
25		$3-HOOC-C_6H_4-$
26		$4-HOOC-C_6H_4-$
27		1-Naphthyl
28		2-Naphthyl
29		$5-Me_2N-1-$Naphthyl
30		2-Thienyl

Cont…

Comp.	Parent structure of comp.	R
31		4 F-C_6H_4-
32		4Cl-C_6H_4-
33		4 CH_3-C_6H_4-
34		2 CH_3 -C_6H_4-
35		4 F-C_6H_4-
36		3 Cl-C_6H_4-
37		4 Cl-C_6H_4-
38		2,4-F_2-C_6H_3-
39		3,4-$Cl_2C_6H_3$-
40		1-Naphthyl
41		4 NO_2-C_6H_4-
42		2 NO_2-C_6H_4-
43		2,4 $(NO_2)_2$-C_6H_3-
44		

The statistical parameters obtained are very useful to investigate the participation of each of the descriptors for modeling the activity. The modeling will be effectively carried out using softwares: Regress-1 (36), martha (39), origin (40) and NCSS. Finally, the proposed QSAR models will be cross validated by leave-on-out procedure.

RESULTS AND DISCUSSION

The results obtained in the present study for modeling CLOHI collagenase inhibitiors on the basis of mainly used Randic, Kier-Hall and MRI as the main correlating parameter. A series of 44 CLOHI collagenase inhibitors are used for this purpose. The modeling is carried out using correlation analysis using the method of least squares (the correlating parameters in terms of A, B, C, D,…etc, and coding C1, C2, C3..). A variety of models were obtained and their statistical significance as well as predictive power were judged. The variety of topological indices were calculated for this set using DRAGON software. From this large set of descriptors, the useful descriptors are selected using the methodology of variable selection in multiple regression analysis. Finally, the models were validated using valadition technique. Several cross-validated parameters were used for this purpose. The most appropriate models were then discussed in Table 4.

Table 2 :

Comp.	log K_i	^0X	^1X	^2X	^0X_v	^1X_v	^2X_v	MRI
1	1.88	16.21	9.66	9.48	12.2	7.65	6.32	9.11124
2	1.88	18.11	10.91	11.08	12.84	7.61	6.45	7.561476
3	1.72	18.11	10.91	11.08	15.1	8.14	9.56	7.561476
4	1.04	26.21	14.66	15.96	16.6	9.49	8.22	9.325125
5	0.85	33.11	18.41	20.83	20.37	11.38	9.94	17.07181
6	1.84	11.78	10.61	10.09	13.65	7.87	7.01	12.48185
7	1.73	19.32	12.27	11.27	14.59	8.98	7.33	15.17126
8	1.7	20.03	12.74	11.81	15.3	9.56	7.78	16.34983
9	1.54	20.19	12.66	11.89	14.89	9.05	7.47	13.33731
10	1.51	20.19	12.66	11.89	15.65	9.43	7.91	13.33731
11	1.48	20.19	12.66	11.89	16.48	9.85	8.39	13.33731
12	1.57	20.19	12.66	11.89	17.05	10.13	8.72	13.33731
13	1.56	20.19	12.66	11.89	15.51	9.36	7.83	20.39021
14	1	21.77	13.57	12.79	15.78	9.45	7.77	15.62115
15	1.08	21.77	13.57	12.8	15.78	9.45	7.77	15.62115
16	1.11	21.77	13.59	12.73	15.78	9.46	7.71	15.62115

Cont…

Comp.	log K_i	^0X	^1X	^2X	^0X_v	^1X_v	^2X_v	MRI
17	0.95	22.64	13.98	13.33	16.83	9.94	8.29	14.82948
18	1.04	23.18	14.56	13.62	17.33	10.23	8.24	17.4597
19	0.95	22.47	14.06	13.28	16.83	9.98	8.08	16.92891
20	0.78	22.69	13.87	13.86	16.15	9.68	8.04	14.9419
21	1.11	21.06	13.07	12.44	16.7	9.91	8.4	13.90384
22	1.32	20.9	13.2	12.06	15.92	9.48	7.69	16.29979
23	1.23	21.93	13.48	12.99	17.36	10.2	8.68	44.35112
24	1.08	21.93	13.48	11.95	17.07	10.05	8.56	13.79171
25	0.95	21.77	13.57	12.8	15.87	9.54	7.85	15.65892
26	0.78	21.77	13.57	12.8	15.87	9.54	7.85	15.65892
27	0.85	21.89	14.25	13.18	16.74	10.36	8.46	20.09206
28	0.9	21.89	14.25	13.18	16.74	10.36	8.46	20.09206
29	0.85	24.34	15.57	14.64	19.11	11.4	9.51	21.85441
30	1.04	18.61	11.77	10.91	14.66	9.32	8.12	11.9554
31	0.95	22.47	14.05	13.3	16.3	9.8	7.75	15.44368
32	0.85	22.47	14.05	13.3	17.05	10.17	8.19	15.44368
33	1.08	22.47	14.05	13.3	16.92	10.11	8.11	17.46
34	0.9	22.47	14.06	13.21	16.92	10.11	8.03	17.46
35	1.08	19.97	12.81	11.51	14.26	7.82	5.64	15.34674
36	1.15	19.97	12.81	11.52	15.01	8.2	6.08	15.34674
37	1.26	19.97	12.81	11.51	15.01	8.2	6.08	15.34674
38	1.18	20.85	13.22	12.04	14.56	7.93	5.76	14.91671
39	1.11	20.85	13.22	12.02	16.07	8.68	6.58	14.91671
40	1	21.67	14.4	12.78	16.11	9.13	6.66	22.87777
41	1.04	19.97	12.81	11.53	14.96	8.45	6.62	15.44206
42	0.85	19.97	12.83	11.45	14.96	8.46	6.56	15.44206
43	0.95	22.42	14.13	12.99	16.14	8.96	7	16.33383
44	0.78	20.68	13.33	11.76	15.68	8.58	6.22	15.98701

Where,

C_2 = log K_i = Inhibitory activity

A C_3 = ^0X = Zero order Randic connectivity index

B C_4 = ^1X = First order Randic connectivity index

C C_5 = ^2X = order Randic connectivity index

D C_6 = $^0X^v$ = Kier-Hall Valence Zero connectivity index

E C_7 = $^1X^v$ = Kier-Hall Valence First order connectivity index

F C_8 = $^2X^v$ = Kier-Hall Valence Second order connectivity index

G C_9 = MRI = Molecular redundancy index

Table 3: Selection results section

Model Size	R-Squared	R-Squared change	Coded variables
1	0.535660	0.535660	B
2	0.639986	0.104326	BC
3	0.651465	0.011479	BCD
4	0.655768	0.004303	BCDG
5	0.658783	0.003015	ABCDG
6	0.659928	0.001145	ABCDEF
7	0.662040	0.002112	ABCDEFG

In order to know the significant model, a graph between R^2 and variable count (x axis) were plotted. These curves become parallel to the X axis, when the number of descriptors 4 indicating that the maximum of 4 descriptors can be used for modeling log K_i.

$$\log K_i = 3.76058672153128 - .406144579481322*C4 + .162802573614171*C5 + 4.36441855459236E\text{-}02*C6 + 4.65815466383864E\text{-}03*C9 \quad ...(2)$$

N = 44, R^2 = 0.6558, R^2A = 0.6205, CV = 0.1717 and F = 18.574

For further examination and statistical preference of the model, Ridge regression analysis has been carried out for the parameters involved in the above model.

Table 4: Ridge regression report (4, 5, 6, 9, mod-4) correlation matrix section

	C4	C5	C6	C9	C2
C4	1.000000	0.917332	0.881570	0.374941	-0.731888
C5	0.917332	1.000000	0.819350	0.227017	-0.542793
C6	0.881570	0.819350	1.000000	0.429055	-0.586069
C9	0.374941	0.227017	0.429055	1.000000	-0.289344
C2	-0.731888	-0.542793	-0.586069	-0.289344	1.000000

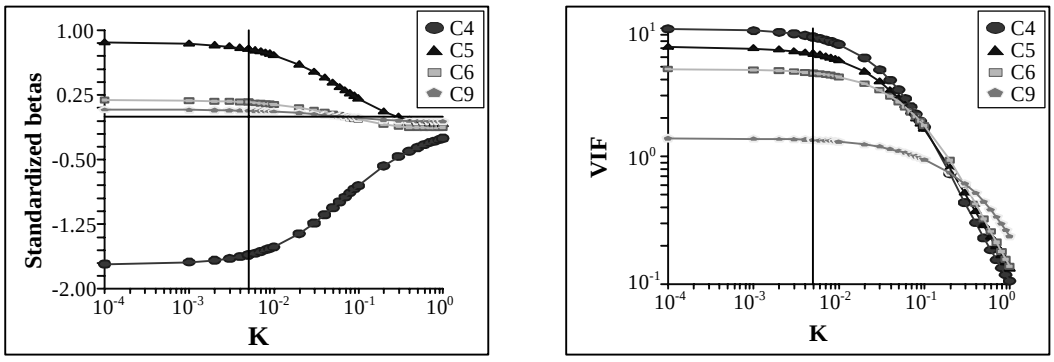
Table 5: Least squares multicollinearity section

Independent variable	Variance inflation	R-Squared vs. other X's	Tolerance
C4	9.9368	0.8994	0.1006
C5	7.1502	0.8601	0.1399
C6	4.8197	0.7925	0.2075
C9	1.3846	0.2778	0.7222
Since all VIF's are less than 10, multicollinearity is not a problem			

Table 6: Eigenvalues of correlations

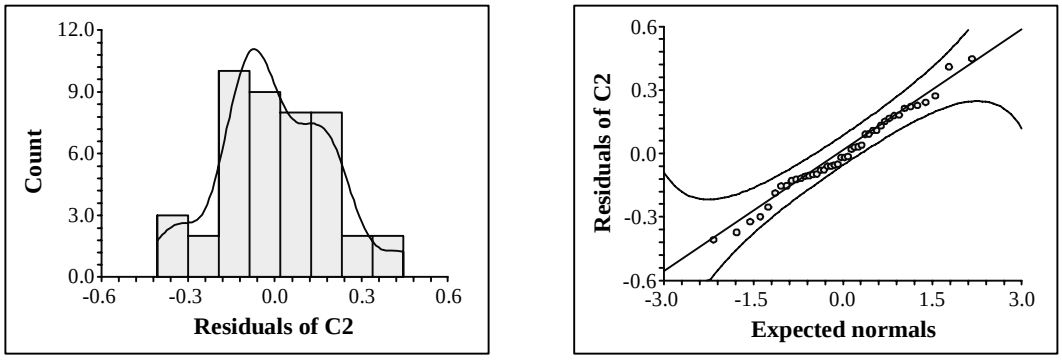
Eigenvalue	Incremental percent	Cumulative percent	Condition number
2.930240	73.26	73.26	1.00
0.846656	21.17	94.42	3.46
0.158405	3.96	98.38	18.50
0.064699	1.62	100.00	45.29
All condition numbers are less than 100. Multicollinearity is not a problem			

In this paper, we discuss the results with a view to obtain statistically significant models. Then by performing regression analysis, we proposed the statistically most significant models for modeling the referred activity. Finally, we have also discussed presence of any statistical draw back in the proposed model. The correlating parameters in terms of capital alphabets : A, B, C, D, E,… etc. and coding than as C_1, C_2, C_3… etc. finally giving the meaning attached to A, B, C… and thus, C_1, C_2, C_3… etc.



Ridge trace **Variance inflation factor plot**

Fig. 1: Ridge trace section



Histogram of residuals of C2 **Normal probability plot of residuals of C2**

Fig. 2: Residual plots section

We now discuss the aforementioned inhibitory activity of 44 compounds. The parameters used are as follows :

C_2 = log K_i = Inhibitory activity

A $C_3 = {}^0X$

B $C_4 = {}^1X$

C $C_5 = {}^2X$

D $C_6 = {}^0X^v$

E $C_7 = {}^1X^v$

F $C_8 = {}^2X^v$

G C_9 = MRI

The variable selection for multiple regression analysis indicated possible 7 regression models (Table 3). A plot of R^2 vs number of variable gave inflection at 4 indicating that we can safely use four variable for modeling log k_i.

(i) One variable model for modeling log k_i

The one variable model for modeling log k_i contain C_4 (^1X) as the correlating parameters and the model is found as below :

Log k_i = 3.4789 – 0.1735 (± 0.0249) C_4

N = 44, R^2 = 0.5357, R^2 A = 0.5248, CV = 0.1521 and F = 48.451

The negative coefficient of C_4 (^1X) indicates that decrease in first order branching increases exhibition of log k_i.

(ii) Two variable model for modeling log k_i

Addition C_5 (^2X) for the above model results into two parameters model with improved statistics. The model is found as below :

log k_i = 3.8870 – 0.3499 (± 0.0558) C_4 + 0.1544 (± 0.0448) C_5 …(4)

N = 44, R^2 = 0.6400, R^2A = 0.6224, CV = 0.1714 and F = 36.442

The opposite signs of C_4 and C_5 are due to their mutual correlatedness. The model once again indicates that decrease in higher order branching increase log k_i.

(iii) Three variable model for modeling log k_i

Further addition of C_6 ($^0X^+$) indicates slight improvement in the statistics as below -

log k_i = 3.6814-0.3941 (±0.0676) C_4+ 0.1515 (±0.0447) C_5 + 0.0521 (±0.0454) C_6 …(5)

N = 44, R^2 = 0.6515, R^2A = 0.6253, CV = 0.1706 and F = 24.922

Once again the opposite signs of the coefficient C_5 and C_6 is due to their inter-correlatedness. This model once again shows that increases in branches decreases the exhibition of log k_i.

(iv) Four variable modeling for log k_i

This and still higher parameters models although show increase in the magnitude of R^2 but all of them contain one or more correlating parameters, whose standard deviations are

much more higher that the respective coefficients and that such models are not allowed statistically.

ACKNOWLEDGEMENT

The authors are thankful to Late Dr. P. V. Khadikar for providing necessary guidance.

REFERENCES

1. C. T. Supuran and A. Scozzoferva, Exp. Opin. Theripatent., **12**, 217 (2002).
2. C. T. Supuran and A. Casini, Med. Res. Rev., **23**, 146 (2003).
3. K. S. Smith and J. G. Ferry, FEMS Microbiol. Rev., **24**, 335 (2000).
4. C. T. Supuran and A. Scozafava, Cur. Med. Chem. Imm. Endoc. Metdls, Agents, **1**, 61 (2001).
5. C. T. Supuran and A. Scozafava, Exp. Opin. Ther. Patents, **10**, 575 (2000).
6. M. A. Hies, B. Masereel, S. Rolin, A. Scozzafeva, G. Calnpeanu, V. Cimpeanu, C. T. Supuran, Bioorg. Med. Chem., **12**, 2717 (2004).
7. C. T. Supuran A. Scozzafava, L. Menabuori, F. Mincion and C. T. Brigants, Eu. J. Pharm. Sci., **8**, 317 (1999).
8. B. E. Mattioni, P. C. Jurs, J. Chem. Inf. Comput. Sci., **42**, 94 (2002).
9. P. V. Khadikar, O. Deeb. A. Jaber, J. Singh, V. K. Agrawal, S. Singh and M. Lakhwani, Lett. Drug Design Discovery (In Press).
10. A. T. Balaban, S. Basak, A. Beteringhe, D. Mills and C. T. Supuran, Mol. Discovery, **8**, 401 (2004).
11. A. T. Balaban, P. V. Khadikar, C. T. Supuran, A. Thakur and M. Thakur, Bioorg. Med. Chem. Lett., **15**, 3966 (2005).
12. J. Singh, M. Lakhwani, P. V. Khadikar, A. T. Balaban, B. W. Clare and C. T. Supuran, Rev. Roum Chem. (In Press).
13. P. V. Khadikar, B. W. Clare, A. T. Balabar, K. T. Supuran, V. K. Agrawal, S. Singh, A. Joshi and M. Lakhwani, Rev. Roum Chem. (In Press).
14. P. V. Khadikar, (Unpublished results).

15. A. Thakur, M. Thakur, P. V. Khadikar and C. T. Supuran, Bioorg. Med. Chem., **12**, 789 (2004).

16. V. K. Agrawal and P. V. Khadikar, Bioorg. Med. Chem. Lett.,**13**, 447 (2003).

17. M. Jaiswal, P. V. Khadikar and T. Supuran, Bioorg. Med. Chem., **12**, 2477 (2004).

18. M. Jaiswal, P. V. Khadikar, A. Scozzafava and C. T. Supuran, Bioorg. Med. Chem. Lett., **14**, 3283 (2004).

19. V. K. Agrawal, S. Bano, C. T. Supuran and P. V. Khadikar, Eu. J. Med. Chem., **39**, 593 (2004).

20. P. V. Khadikar and D. J. Mandloi, Bioinformoics, India, **2**, 86 (2004).

21. P. V. Khadikar, V. Sharma, S. Karmarkar and C. T. Supuran, Bioorg. Med. Chem. Lett., **15**, 931 (2005).

22. P. V. Khadikar, V. Sharma, S. Karmakar and C. T. Supuran, Bioorg. Med. Chem. Lett., 15, 923 (2005).

23. V. K. Agrawal, S. Singh, P. V. Khadikar and C. T. Supuran, Bioorgvled. Chem. Lett., **16**, 2044 (2006).

24. W. D. Gray, T. H. Maren, G. M. Sisson and F. H. Smith, J. Pharmacol. Exp. Jher, **121**, 160 (1957).

25. M. W. Duffel, I. S. Ing, T. M. Segarra, J. A. Dixon, C. F. Barfknechl and R. D. Schoehwold, J. Med. Chem., 29, 1488 (1986).

26. J. A. Verpoorte, S. I. Mehta and J. T. Edsau, J. Biol. Chem., **242**, 1421 (1967).

27. T. H. Maren and J. Glaucoma, in; W. R. Chegwidden, N. D. Carter, Y. H. Edwans (Eds.), The Carbonic Annydoases; New Horizons, 4, Birkhanser Verlag, Basel (1995) pp. 49-62.

28. C. T. Supuran, A. Scozzafava, L. Menabuoni, F. Mincionl, F. Briganti and C. T. Minicionl, Eu J. Pharm Sci., **8**, 317 (1999).

29. H. Gao and J. Bajorath, Mol. Diversity, **4**, 115 (1999).

30. B. W. Clare and C. T. Supuran, Bioors. Med. Chem., **13**, 2197 (2005).

31. D. Bonehe and D. H. Rouvray, Chemical Topology, Applications and Techniques, Gordon & Breack, New York (2000).

32. D. Bonchev and D. H. Rouvray, Chemical Topology: Introduction and Fundamentals, Gorden & Breach, New York (1999).

33. M. V. Diudea and O. Ivanciace, Molecular Topology, Complex, Cluj (1995).

34. N. Trinajstis, Chemical Graph Theory, CRC Press, Boca Raton (FL) (1993).

35. A. T. Balaban, Chemical Applications of Graph Theorm, Academic Press, London (1976).

36. DRAGON

37. LUKO-1 & REGRESS-1

38. ACD Labs

39. MARTHA

40. ORIGIN

41. S. Charterjee, A. S. Hadi and B. Price, Regression Analysis by Examples, 3rd Ed, Wiley, New York (2000).

42. M. V. Diudea, M. S. Florescu and P. V. Khadikar, Molecular Topology and Its Applications, EFICON, Bucharest (2006) (Eficol Press, Bucuresti)

43. D. Mandloi, QSAR Study on Anti-microbial Activity of Mannicl Bases, Ph.D. Thesis, D. A. V. V. Indore (2004).

44. M. Sharma, Topological Studies on the effects of Hydroxamic acids on DNA Synthesis by Ehrlion Ascites Turner Cells, Ph.D. Thesis, D. A. V. V. Indore (2005).

45. G. S. Thaklral, Ph. D. Thesis, D. A. V. V. Indore (2006).

46. H. J. Wiener, J. Am. Chem. Soc., **69**, 17 (1947).

47. I. Gutman, Graph Theory Notes New York, 27 (1994) pp. 9-15.

48. P. V. Khadikar, N. V. Deshpande, P. P. Kale, A. Dobrynin, I. Gutman and G. Domotor, J. Chem. Inf. Comput. Sci., **35**, 547 (1995).

49. P. V. Khadikar, P. P. Kale, N. V. Deshpande, S. Karmarkar and V. K. Agrawal, Commun. Math. Comput. Chem. (MASCH), **43**, 7 (2001).

50. P. V. Khadikar, S. Karmarkar, V. K. Agrawal, S. Singh, A. Shrivastava, I. Lukovits and M. V. Diudea, Lett. Drug Design Discovery, **2**, 606 (2005).

51. P. V. Khadikar, Nat. Acad. Sci. Lett., **23**, 113 (2000).

52. P. V. Khadikar, S. Karmankar and V. K. Agrawal, J. Chem. Inf. Comput. Sci., **41**, 934 (2001).

53. P. V. Khadikar, N. V. Deshpande, P. P. Kale, S. Karmarkar and V. K. Agrawal, J. Math Chem., **29**, 143 (2001).

54. P. V. Khadikar, M. V. Diudea, J. Singh, P. E. John, A. Shrivastava, S. Singh, S. Karmakar, M. Lakhwani and P. Thakur, Curr. Bioact. Comp., **2**, 19 (2006).

55. P. V. Khadikar, V. K. Agrawal and S. Karmarkar, Bioorg. Med. Chem., 10, 3499 (2002).

Accepted : 21.04.2013