S. Bajaj[1], *S. S. Sambi*[1], *and A. K. Madan*[2]*

TOPOCHEMICAL MODELS FOR PREDICTING THE ACTIVITY OF α,γ-DIKETO ACIDS AS INHIBITORS OF THE HEPATITIS C VIRUS NS5B RNA-DEPENDENT RNA POLYMERASE

[1] School of Chemical Technology, GGS Indraprastha University, Delhi 110 006, India
[2] Faculty of Pharmaceutical Sciences, MD University, Rohtak 124 001, India
* e-mail: madan ak@yahoo.com

A relationship between the topochemical indices of α,γ-diketo acids and their inhibitory activity with respect to the hepatitis C virus NS5b RNA-dependent RNA polymerase has been studied. The values of Wiener's topochemical index W^c (a distance based descriptor), the Zagreb topochemical index M_2^c (an adjacency based descriptor), and the topochemical eccentric connectivity index ξ_c^c (an adjacency-cum-distance based descriptor) were calculated for a set of 30 compounds using an in-house computer program. The resulting data array was analyzed and the models of activity were developed after determination of the active ranges of parameters. Subsequently, a biological activity was assigned using these models to each compound included in the data bank, and the assignment was compared with data reported on the inhibitory activity. A high accuracy of prediction was observed for the proposed models. These models possess vast potential in providing basis structures for the development of diketo acids capable of effectively inhibiting the hepatitis C virus NS5b RNA-dependent RNA polymerase.

INTRODUCTION

Hepatitis C virus (HCV) infection is one of the leading causes of liver disorders in the world. It is a common cause of cirrhosis and hepatocellular carcinoma, as well as the most common reason for liver transplantation. Thus, appropriate therapeutic approaches have a strong clinical impact on the morbidity and mortality of HCV-infected patients [1]. A complex therapy combining interferon-alpha and ribavirin administration represents the current standard treatment for chronic HCV infection, although it has demonstrated limited success and causes some serious side effects. Promising alternative approaches toward the control of HCV infection and the development of new antiviral agents include the use of NS3/4A serine protease and NS5B polymerase inhibitors. Preclinical results from the development of HCV polymerase inhibitors, both nucleoside and non-nucleoside, are promising [2]. The RNA-dependent RNA polymerase of HCV is required for viral RNA replication and thus represents an attractive drug discovery target [3].

All quantitative structure-activity relationship (QSAR) studies for rational drug design are based on the notion that biological activity is the function of physicochemical parameters, structural descriptors, and topological and/or electronic properties of a given molecule [4]. Topological methods are probably the simplest ones because the parameters are easily calculable from the graphical representation of molecules and do not require estimation of any physicochemical property. These methods consider the arrangement of atoms over the skeleton and use the concepts of steric relations and molecular bulk, branching, and relationships of various nonbonded parts of the molecule [5, 6]. In recent years, a large number of topological indices have been proposed and utilized for chemical documentation, isomer discrimination, study of molecular complexity, chirality, similarity/dissimilarity, QSAR/QSPR, drug design, database selection, lead optimization, and rational combinatorial library design and for deriving multilinear regression models [7 – 10]. However, among the reported topological indices, only a few have been successfully employed in QSAR studies. Noteworthy quantities are Wiener's index [11, 12], Hosoya's index [13], Randić's molecular connectivity index χ [14], high-order connectivity indices ${}^n\chi$ for the paths of length n defined by Kier and Hall [15], Balaban's index J [16], Zagreb group parameters M_1 and M_2 [17], eccentric connectivity index [18], and eccentric adjacency index [19]. Topochemical indices are the topological parameters derived from the corresponding chemical graphs weighted with respect to the relative atomic weights of the noncarbon atom present in the given molecule. Previously, we have studied various QSARs using topochemical indices, including the atomic molecular connectivity index [20], topochemical eccentric adjacency index [21], Wiener's topochemical index [22, 23], topochemical superadjacency index [24, 25], topochemical Zagreb indices M_1^c and M_2^c [26], and topochemical eccentric connectivity index [27].

In the present study, three topochemical indices – Wiener's topochemical index W^c (a distance based descriptor), Zagreb topochemical index M_2^c (an adjacency based descriptor), and topochemical eccentric connectivity index ξ_c^c (an adjacency-cum-distance based descriptor) – have been used to develop models for predicting the inhibitory activity of α,γ-diketo acids with respect to hepatitis C virus NS5b RNA-dependent RNA polymerase.

CALCULATION OF TOPOCHEMICAL INDICES

Wiener's topochemical index W^c is a modification of the oldest and most widely used distance-based topological parameter (conventional Wiener's index) [11 – 12]. This modified index, which takes into consideration the presence and relative position of heteroatom(s) in a molecular structure, is defined as the sum of the chemical dis-

tances between all pairs of vertices in hydrogen suppressed molecular graph G [22]:

$$W^c = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} P_{ij}^c, \quad (1)$$

where, P_{ij}^c is the "chemical" length of the path that contains the least number of edges between vertex *i* and vertex *j* in the graph G; *n* is the maximum possible value of *i* and *j* (the number of vertices in the graph G). Wiener's topochemical index (W^c) can be easily calculated using the chemical distance matrix of a hydrogen suppressed molecular structure. This matrix is obtained by substituting row elements corresponding to heteroatom(s), with relative atomic weights with respect to carbon atom [22, 23].

The Zagreb topochemical index M_2^c is a modification of the Zagreb group parameter (or Zagreb index) M_2, which was introduced by Gutman and Trinajstić [17]. This quantity is also sensitive to the presence and the arrangement of heteroatoms in a given molecule. The modified index has been reported to have lower degeneracy in comparison to the original value [26]. It is defined as a sum of the chemical weights of all edges in hydrogen suppressed molecular graph G:

$$M_2^c(\mathrm{G}) = \sum_{ij}^{n} d^c(i)d^c(j), \quad (2)$$

where $d^c(i)d^c(j)$ is the chemical weight of the $\{i, j\}$ edge in hydrogen suppressed molecular graph G and *n* is the number of edges [26].

The Zagreb topochemical index M_2^c can be easily calculated from the chemical adjacency matrix of hydrogen suppressed molecular structure. The chemical degree of the *i* th vertex is the sum of entries in the *i* th row of the chemical adjacency matrix. The adjacency matrix weighted with respect to the heteroatom present within the molecule is called the chemical adjacency matrix. This matrix is obtained by substituting row elements, corresponding to heteroatom, with relative atomic weights with respect to carbon atom. Thus, in this matrix, nonzero row elements represent the chemical adjacency between the corresponding vertices in a molecular graph.

TABLE 1. Relationship between Topochemical Indices and Anti-HCV Polymerase Activity of α,γ-Diketo Acids

Compound	Basic structure (Fig. 1)	Topochemical index	Anti-HCV activity						
						Assigned			Reported
			W^c	M_2^c	ξ_c^c	W^c	M_2^c	ξ_c^c	
1	I	Ph	333.67	77.55	185.76	–	–	–	–
2	I	Me	101.33	42.882	73.77	–	–	–	–
3	I	*t*-Bu	214.33	64.88	115.43	–	–	–	–
4	I	$(Ph)_2CH$	925.32	123.55	329.98	–	–	–	–
5	I	3-thiophene	281.99	88.55	201.33	–	–	–	–
6	I	3-pyridine	343.01	81.72	194.84	–	–	–	–
7	I	4-pyridine	337.67	80.78	192.72	–	–	–	–
8	I	2-furan	275.99	77.32	166.09	–	–	–	–
9	II	Me	411.65	83.88	223.54	–	–	–	–
10	IIA	H	333.66	77.55	185.76	–	–	–	–
11	IIB	H	285.99	67.55	174.66	–	–	–	–
12	III	4-Me	407.32	84.55	218.32	–	–	–	–
13	III	4-Et	495.99	89.55	254.32	–	–	–	–
14	III	4-*t*-Bu	679.32	105.55	293.65	–	–	–	–
15	III	2-O-propyl	675.14	102.21	295.75	–	–	–	–
16	III	2-O-butyl	803.63	106.21	341.41	–	–	–	–
17	III	2-O-pentyl	951.13	110.21	389.07	–	–	–	–
18	III	2-O-Bn	1203.12	130.54	456.07	–	–	–	–
19	III	$2\text{-}OCH_2CH_2CH_2CN$	952.71	110.71	396.14	–	–	–	+
20	III	3-NHCOPh	1372.86	137.55	509.69	–	–	–	–
21	III	$3\text{-}NHSO_2Ph$	1739.82	198.55	677.41	+	+	+	–
22	III	3-NH-Bn	1249.11	127.38	486.52	–	–	–	–
23	III	3-O-Bn	1267.12	129.21	496.07	–	–	–	–
24	III	$3\text{-}CH_2OPh$	1264.45	129.21	496.07	–	–	–	–
25	IV	2-CNPh	1589.03	142.88	553.08	+	+	+	+
26	IV	3-CNPh	1621.03	141.88	583.47	+	–	+	–
27	IV	CH_2CH_2CN	691.74	98.05	316.60	–	–	–	–
28	IV	2-CN-5-BrPh	1832.61	178.21	820.08	+	+	+	+
29	IV	2-CN-3-ClPh	1784.10	162.63	654.69	+	+	+	+
30	IV	$2\text{-CN-3,5-}Cl_2Ph$	1989.13	179.42	705.82	+	+	+	+

Note: (+) Active compound (IC_{50} < 1 μM); (–) inactive compound.

The eccentric connectivity index ξ_c^c was originally introduced by Sharma, et al. [18]. A topochemical extension of this quantity has been recently introduced by Kumar et al. [27]. This modified index is defined as the sum of the products of the chemical eccentricity and chemical degree of each vertex in hydrogen suppressed molecular graph G:

$$\xi_c^c = \sum_{i=1}^{n} (E_i^c V_i^c), \qquad (3)$$

where E_c^i is the chemical eccentricity and V_c^i is the chemical degree of *i* th vertex in graph G and *n* is the number of vertices.

For a molecular graph G containing *n* vertices (v_1, v_2, ..., v_n), the number of first neighbors of the *i* th vertex is called the chemical degree of this vertex and denoted by deg(v_{ic}). The chemical distance d^c (v_i, v_j|G) between the vertices v_i and v_j of graph G is the length of the shortest path connecting v_i to v_j. The chemical eccentricity E_i^c of vertex v_i in graph G is the length of the shortest path from v_i to the vertex v_j that is farthest from v_i (E_{ic} = max[d^c(v_i, v_j)]; j| G). The topochemical eccentric connectivity index is calculated from the chemical distance matrix (D^c) and the chemical adjacency matrix (A^c), whch are used for determining the chemical eccentricity and the chemical degree of vertices, respectively. The topochemical eccentric connectivity index was found to exhibit much lower degeneracy than the conventional eccentric connectivity index and is also sensitive to the presence and arrangement of heteroatoms without compromising with the discriminating power of the original index [27].

CONSTRUCTION OF TOPOCHEMICAL MODELS

A set of data for 30 α,γ-diketo acids [28] exhibiting inhibitory activity with respect to Hepatitis C virus NS5b RNA-dependent RNA polymerase was selected for the present study. The basic structures of these compounds are presented in Figure 1. The values of Wiener's topochemical index, the Zagreb topochemical index M_2^c and the topochemical eccentric connectivity index for each compound were calculated using an in-house computer program. The resulting data array was analyzed and suitable models were developed after determining the activity ranges by maximization of the moving average with respect to active compounds [19 – 26, 29]. Subsequently, each analog involved in the initial set was assigned a biological activity

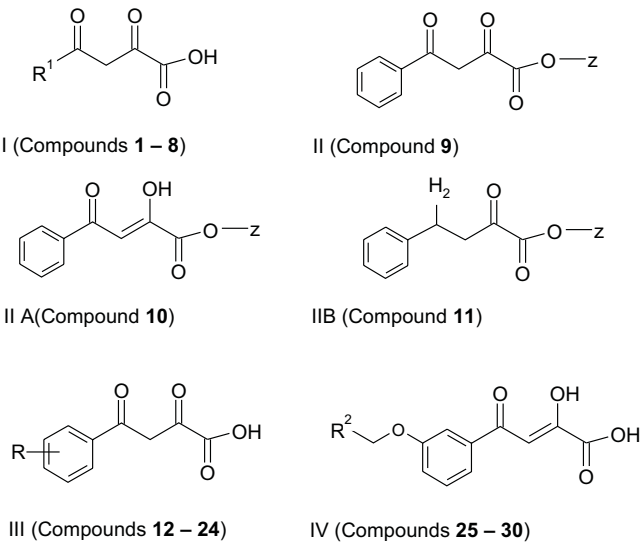

Fig. 1. Basic structures of diketo acids.

in terms of these models. This assignment was then compared with the reported activity in terms of IC_{50} (μM). For the purpose of this study, the analogs having IC_{50} below 1 μM were considered active. The percentage of correct prediction for each range was calculated as a ratio of the number of compounds for which the activity was correctly predicted to the total number of compounds present in the corresponding range. The overall degree of prediction was determined as the ratio of the total number of compounds with correctly predicted activity to the total number of analogs present in both active and inactive ranges. The results are summarized in Tables 1 and 2.

RESULTS

Retrofit analysis of the data in Tables 1 and 2 provides the following information for the models based upon various topochemical indices.

A model based upon Wiener's topochemical index (W^c):

(i) Biological activity was assigned to all 30 compounds involved in the initial data set; the inhibitory activity with respect to Hepatitis C virus NS5b RNA-dependent RNA polymerase was correctly predicted for 27 compounds (90%).

TABLE 2. Models for Prediction of anti-HCV Polymerase Activity of α,γ-Diketo Acids

Model index	Character of model	Index range	Number of compounds		Percentage accuracy	Average IC_{50} (μM)		Overall accuracy of prediction
			Total	Correct		Total	Correct	
W^c	Inactive	< 1589.00	24	23	95.8	29.75	31.01	
	Active	> 1589.00	6	4	66.7	3.86	0.12	90.0
M_c^2	Inactive	< 10.87	25	24	96.0	29.19	30.39	
	Active	> 10.87	5	4	80.0	1.44	0.12	93.3
ξ_c^c	Inactive	< 553.00	24	23	95.8	29.73	31.01	
	Active	> 553.00	6	4	66.7	3.86	0.12	90.0

(ii) The inactive range corresponded to Wiener's topochemical index below 1589.00. The activity was correctly predicted for 23 (95.8% of a total of 24) compounds falling in the inactive range. The IC_{50} (μM) of the correctly predicted compounds in the inactive range was 31.01.

(iii) The active range corresponded to Wiener's topochemical index values greater than 1589.00. The activity was correctly predicted for 4 (66.67% of a total of 6) compounds falling in the active range. The average IC_{50} (μM) of the correctly predicted compounds was 0.12.

(iv) The following expression was obtained for estimation of the average IC_{50} (μM) within the framework of this model:

$$IC_{50}\ (\mu M)^{calcd} = 0.00151(W^c) + 3.3251. \qquad (4)$$

A model based upon the Zagreb topochemical index (M_2^c):

(i) Biological activity was assigned to all 30 compounds involved in the initial data set; the inhibitory activity with respect to Hepatitis C virus NS5b RNA-dependent RNA polymerase was correctly predicted for 28 compounds (93.3%).

(ii) The inactive range corresponded to Wiener's topochemical index below 142.87. The activity was correctly predicted for 24 (96.0% of a total of 25) compounds falling in the inactive range. The IC_{50} (μM) of the correctly predicted compounds in the inactive range was 30.0.

(iii) The active range corresponded to Wiener's topochemical index values greater than 142.87. The activity was correctly predicted for 4 (80% of a total of 5) compounds falling in the active range. The average IC_{50} (μM) of the correctly predicted compounds was 0.12.

(iv) The following expression was obtained for estimation of the average IC_{50} (μM) within the framework of this model:

$$IC_{50}(\mu M)^{calcd} = 0.016(M_2^c) + 3.27 \qquad (5)$$

A model based upon the topochemical eccentric connectivity index (ξ_c^c):

(i) Biological activity was assigned to all 30 compounds involved in the initial data set; the inhibitory activity with respect to Hepatitis C virus NS5b RNA-dependent RNA polymerase was correctly predicted for 27 compounds (90%).

(ii) The inactive range corresponded to Wiener's topochemical index below 553.00. The activity was correctly predicted for 23 (95.8% of a total of 24) compounds falling in the inactive range. The IC_{50}(μM) of the correctly predicted compounds in the inactive range was 31.01.

(iii) The active range corresponded to Wiener's topochemical index values greater than 553.00. The activity was correctly predicted for 4 (66.67% of a total of 6) compounds falling in the active range. The average IC_{50} (μM) of the correctly predicted compounds was 0.12.

(iv) The following expression was obtained for estimation of the average IC_{50} (μM) within the framework of this model:

$$IC_{50}(\mu M)^{calcd} = 0.003(\xi_c^c) + 2.6143. \qquad (6)$$

DISCUSSION

A current trend in theoretical chemistry, molecular pharmacology, toxicology and environmental chemistry is the prediction of properties of molecules using their physicochemical characteristics and/or theoretical parameters. A major part of the current research in mathematical chemistry, chemical graph theory and QSAR/QSPR studies involve topological indices. Topological indices are numerical invariants that quantitatively characterize molecular structure [30]. A major disadvantage of the topological indices is that they are not sensitive to the presence of heteroatoms in the molecule. The topochemical indices overcome this limitation, since they take into consideration the presence and arrangement of heteroatoms in a molecule. Consequently, the degeneracy is much lower and, hence, topochemical models are considered advantageous in comparison to topological models for predicting the activity of molecules. The idea behind choosing the topochemical indices for the present study was that these three quantities explain the structural properties based on three different concepts, since Wiener's topochemical index is a distance based descriptor, the Zagreb topochemical index M_2^c is an adjacency based quantity, and the topochemical eccentric connectivity index is based on both distance and adjacency.

The NS3 protease and NS5b polymerase are the most studied targets for anti-HCV therapy. The α,γ-diketo acids has been identified as low-micro-molar, specific, and reversible inhibitor of HCV NS5b polymerase. Their mechanism of action was reported to be different from that of all other known HCV polymerase inhibitors, since they are capable of interacting directly with metal ions present in the active site of HCV NS5b polymerase. Diketo acids have been reported as potent HIV integrase and RNase-H inhibitors, with anti-HIV activity in cell based assays. Similar inhibitors of influenza virus are also known [28]. Diketo acids have been recently also identified as inhibitors of HCV NS5b polymerase and have the potential to be developed further.

In conclusion, topochemical models have been developed for predicting the anti-HCV activity of diketo acids. The attractive feature of these models is the high potency in the active ranges, as evidenced by the fact that the average IC_{50} (μM) of correctly predicted compounds in the active range is only 0.12. Exceptionally high accuracy of prediction of the proposed models offer vast potential for providing basic structures for the development of potent therapeutic agents.

REFERENCES

1. S. Fargion, A. L. Fracanzani, and L. Valenti, *J. Antimicrob. Chemother.*, **53**(5), 708 – 712 (2004).
2. Z. J. Ni and A. S. Wagman, *Curr. Opin. Drug Discov. Devel.*, **7**(4):446 – 459 (2004).
3. R. T. Sarisky, *J. Antimicrob. Chemother.*, **54**(1), 14 – 16 (2004).
4. H. Kuainyi, in: *Burger's Medicinal Chemistry and Drug Discovery*, M. E. Wolf (ed.), Wiley, New York (1995), Vol. 1, pp. 497.

5. N. Trinajstić: in: *Chemical Graph Theory*, CRC, Boca Raton, FL (1992).
6. A. Saha, K. Roy, K. De, and C. Sengupta, *J. Ind. Chem. Soc.,* **78**, 92 – 97 (2001).
7. S. Nikolić, G. Kovačević, A. Miličević, and N. Trinajstić, *Crat. Chem. Acta*, **76**(2), 113 – 124 (2003).
8. E. Estrada and E. Uriarte, *Curr. Med. Chem.*, **8**(13),1573 – 1588 (2001).
9. R. Gozalbes, J. Doucet, and F. Derouin, *Curr. Drug Targets Infect. Disord.*, **2**(1), 93 – 102 (2002).
10. E. Estrada, G. Patlewicz, E. Uriate, *Ind. J. Chem.*, 42A, 1315 – 1329 (2003).
11. H. Wiener, *J. Am. Chem. Soc.*, **69**, 17 – 20 (1947).
12. H. Wiener, *J. Am. Chem. Soc.*, **69**, 2636 – 2638 (1947).
13. H. Hosoya, *Bull. Chem. Soc. Jpn.*, **44**, 2332 – 2337 (1971).
14. M. Randić, *J. Am. Chem. Soc.*, **97**, 6609 – 6615 (1975).
15. L. B. Kier, L. H. Hall, W. J. Murray, and M. Randic, *J. Pharm. Sci.*, **64**, 1971 – 1974 (1975).
16. A. T. Balaban, *J. Chem. Inf. Comput. Sci.*, **25**, 334 – 343 (1985).
17. I. Gutman and N. Trinajstic, *Chem. Phys. Lett.*, **17**, 535 – 545 (1973).
18. V. Sharma, R. Goswami, and A. K. Madan, *J. Chem. Inf. Comput. Sci.*, **37**, 273 – 281 (1997).
19. S. Gupta, M. Singh, and A. K. Madan, *J. Computer-Aided Mol. Design.*, **15**, 671 – 678 (2001).
20. S. Gupta, M. Singh, and A. K. Madan, *J. Mol. Struct. (Theochem)*, **571**, 147 – 152 (2001).
21. S. Gupta, M. Singh, and A. K. Madan, *Ind. J. Chem.*, **42A**, 1414 – 1425 (2003).
22. S. Bajaj, S. S. Sambi, and A. K. Madan, *J. Mol. Struct. (Theochem)*, **684**, 197 – 203 (2004).
23. S. Bajaj, S. S. Sambi, and A. K. Madan, *MATCH: Commun. Math. Comput. Chem.*, **55**, 193 – 204 (2006).
24. S. Bajaj, S. S. Sambi, and A. K. Madan, *QSAR & Comb. Sci.*, **23**(7), 506 – 514 (2004).
25. S. Bajaj, S. S. Sambi, and A. K. Madan, *Bioorg. Med. Chem Lett*., **15**(2), 467 – 469 (2005).
26. S. Bajaj, S. S. Sambi, and A. K. Madan, *Crat. Chem. Acta*., **78**(2), 165 – 174 (2005).
27. V. Kumar, S. Sardana, and A. K. Madan, *J. Mol. Model*, **10**, 397 – 407 (2004).
28. V. Summa, A. Petrocchi, P. Pace, et al., *J. Med. Chem.*, **47**, 14 – 17 (2004).
29. S. Bajaj, S. S. Sambi, and A. K. Madan, *Bioorg. Med. Chem.*, **12**, 3695 – 3701 (2004).
30. S. C. Basak, A. T. Balaban, G. D. Grunwald, and B. D. Gute, *J. Chem. Inf. Comput. Sci.*, **40**, 891 – 898 (2000).

Submitted 11.01.05

ТОПОХИМИЧЕСКИЕ МОДЕЛИ ДЛЯ ПРЕДСКАЗАНИЯ АКТИВНОСТИ α,γ-ДИКЕТОКИСЛОТ В ОТНОШЕНИИ РНК-ЗАВИСИМОЙ РНК-ПОЛИМЕРАЗЫ ВИРУСА ГЕПАТИТА С

С. Баджадж, С. С. Самби, А. К. Мадан

Колледж химической технологии, Университет Индрапраста, Дели, Индия;
Фармацевтический Факультет, Медицинский Университет, 124001 Рохтак, Индия

Установлена взаимосвязь между характеристиками структуры α,γ-дикетокислот и их способностью ингибировать активность РНК-зависимой РНК-полимеразы вируса гепатита С. В качестве структурных дескрипторов использованы модифицированные топохимические индексы Винера (W^c, дескриптор межатомных расстояний), Загребской группы (M_2^c, дескриптор соседства), и молекулярной связности (ξ_c^c, комбинированный дескриптор), значения которых были рассчитаны для 30 соединений обучающей выборки. Полученная база данных использована для построения моделей активности на основе каждого из выбранных топохимических индексов. Результаты предсказания активности с использованием построенных моделей с хорошей точностью совпадают с имеющимися экспериментальными данными, что позволяет использовать эти модели в поиске базовых соединений для создания антивирусных препаратов, подавляющих активность РНК-зависимой РНК-полимеразы вируса гепатита С.