Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 467–469
Bioorganic & Medicinal Chemistry Letters

Topochemical model for prediction of anti-HIV activity of HEPT analogs

Sanjay Bajaj,[a] S. S. Sambi[a] and A. K. Madan[b,*]

[a]*School of Chemical Technology, GGS Indraprastha University, Delhi 110006, India*
[b]*Faculty of Pharmaceutical Sciences, MD University, Rohtak 124001, India*

Received 22 August 2004; accepted 12 October 2004
Available online 6 November 2004

Abstract—The relationship between the superadjacency topochemical index and the anti-HIV activity of HEPT analogs has been investigated in the present study. The values of superadjacency topochemical index of all the analogs involved in the data set were calculated using an in-house computer program. Resulting data were analyzed and a suitable model was developed after identification of the active range. Subsequently, a computed biological activity was assigned to each of the compounds involved in the dataset, which was then compared with the reported anti-HIV activity. Accuracy of prediction was found to be 88% using the said model. The predictive ability of the model indicates that this model can be used for predicting the anti-HIV activity of the compounds prior to synthesis and may prove to be highly beneficial for providing lead structures for development of potent anti-HIV agents.

Graph theory is largely applied to the characterization of chemical structures, as well as to qualitative and quantitative structure–property (QSPR) and structure–activity (QSAR) relations by means of certain numerical characteristics, the so-called topological indices.[1–3] In the last decade, topological indices have emerged as powerful tools for predicting biological activity of molecules, designing combinatorial libraries and lead identification[4,5] and form an integral part of new molecular research especially anti-HIV research. This is because anti-HIV therapy today is in need of faster introduction of newer agents to fight the epidemic because each day approximately 14,000 new infections are occurring and AIDS is now the leading cause of death in Sub-Saharan Africa and the fourth-biggest killer globally (http://www.worldbank.org). Amongst the developing anti-HIV agents, 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thyamine (HEPT) derivatives is a very well known class. HEPT interacts with reverse transcriptase (RT) at the allosteric binding pocket common to all non-nucleoside inhibitors. Further, phosphorylation is not an obligate step for its ability to block the function of RT. Therefore this is one of the classes, which have the potential to be developed further as anti-HIV agents. In this study the superadjacency topochemical index,[6] a distance-cum-adjacency based molecular descriptor, has been used for the development of model for prediction of anti-HIV activity of HEPT analogs.

Superadjacency topochemical index, denoted by $\int^{Ac}$, is defined as the sum of the products of the concerned vertex chemical degree and the sum of adjacent vertex chemical degrees divided by the chemical eccentricity of the concerned vertex, over all the vertices in the hydrogen suppressed molecular graph. It is expressed as

$$\int^{Ac}(\mathrm{G}) = \sum_{i=1}^{n} \frac{\deg(v_{ic})^{*}\mathrm{S}_{ic}}{E_{ic}}$$

$$S_{ic} = \sum \deg(v_{jc})$$

where S_{ic} is the sum of chemical degrees of all vertices (v_j), adjacent to vertex i and n is the number of vertices in graph G.

For a molecular graph (G), v_1, $v_2 \ldots v_n$ are vertices, the number of first neighbors of a vertex v_i is the chemical degree of this vertex and is denoted by $\deg(v_{ic})$. The chemical distance dc (v_i, v_j |G) between the vertices v_i and v_j of G is the length of the shortest path connecting v_i to v_j. The chemical eccentricity E_{ic} of vertex v_i in graph G is the length of the shortest path from v_i to vertex v_j, that is, farthest from v_i (E_{ic} = max dc (v_i, v_j); j|G).

* Corresponding author. Tel.: +91 1262 272535/1262 212111/11 24333398; e-mail: madan_ak@yahoo.com

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.10.053

The superadjacency topochemical index is calculated from the chemical distance matrix (D^c), the chemical adjacency matrix (A^c) and a new matrix, the additive chemical adjacency matrix (A^{ac}), obtained by modifying A^c. These matrices are obtained by substituting row elements corresponding to heteroatom like N, O, Cl etc., with relative atomic weight with respect to the carbon atom. The chemical distance matrix is utilized for deriving chemical eccentricity while chemical adjacency matrix is utilized for deriving chemical degree of vertices. When non-zero row elements in chemical adjacency matrix represent the chemical degree of corresponding vertex in a molecular graph, the matrix may be defined as the additive chemical adjacency matrix. This matrix is utilized for deriving S_{ic} for the corresponding vertex.[6]

In this study a dataset[7] comprising of 33 HEPT analogs (Table 1, Fig. 1) was used for the development of model for prediction of anti-HIV activity of HEPT analogs. The values of superadjacency topochemical index were calculated for each of the compounds involved in the dataset, using an in-house computer program. This computer program converts the input in the form of structural information into files for information processing, feature selection and activity recognition. The resulting data were analyzed and suitable model was developed after identification of the active range based on maximization of moving average with respect to the active compounds (<35% = inactive, 35–65% = transitional, 65% = active).[6,8] Subsequently, each of the compounds was assigned a biological activity using the said model, which was then compared with the reported anti-HIV activity. The activity of these compounds has been evaluated as their inhibitory effects on the replication of HIV-1 in human T-4 lymphoblastoid CEM-SS cell lines

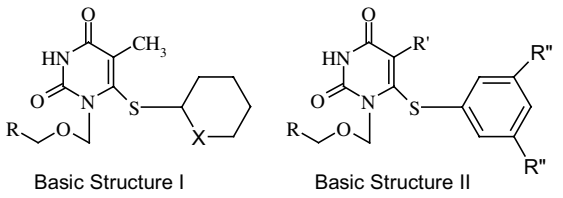

Figure 1. Basic structure of 1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine derivatives.

Table 1. Relationship between *superadjacency topochemical index* and anti-HIV activity of HEPT analogs

Comp. no.	Basic str.	Substituent			$\int^{Ac}$	Anti-HIV activity	
		X	R			Assigned	Reported
1	I	CH	Br		26.46	–	–
2	I	CH	I		23.55	–	–
3	I	CH	Morpholino		32.50	–	–
4	I	CH	Piperazinyl		32.68	–	–
5	I	CH	$N(CH_2CN)_2$		32.13	–	–
6	I	CH	NHPh		30.98	+	–
7	I	CH	$N(Ph)_2$		36.84	–	–
8	I	CH	NH_2		34.70	–	–
9	I	CH	NHCOPh		30.83	–	–
10	I	CH	$N(COPh)_2$		38.47	+	–
11	I	CH	$NHCO(CH_2)_2Cl$		26.93	–	–
12	I	CH	$NHCO(CH_2)PO_3H$		31.40	–	–
13	I	CH	$NHCOCH_2NHCH_2PO_3H$		30.23	–	–
14	I	CH	SPh		30.19	–	–
15	I	CH	SPY		30.37	–	–
16	I	N	NHPh		31.20	–	–
17	I	N	SPh		30.39	–	–
18	I	N	SPy		30.56	–	–
		R′	R	R″			
19	II	CH_3	Phenyl	H	34.97	–	–
20	II	C_2H_5	Phenyl	H	35.97	–	+
21	II	C_2H_5	Phenyl	CH_3	37.88	+	+
22	II	CH_3	Cyclohexyl	H	34.97	–	–
23	II	CH_3	3-Pyridyl	H	35.19	–	–
24	II	CH_3	2-Furyl	H	36.29	–	–
25	II	C_2H_5	2-Furyl	H	37.36	+	+
26	II	C_2H_5	2-Furyl	CH_3	39.46	+	+
27	II	CH_3	2-Thienyl	H	37.01	–	–
28	II	C_2H_5	2-Thienyl	H	38.01	+	+
29	II	C_2H_5	2-Thienyl	CH_3	40.01	+	+
30	II	CH_3	5-Nitro-2-thienyl	H	35.75	–	–
31	II	C_2H_5	5-Nitro-2-thienyl	H	36.55	–	–
32	II	CH_3	2-Benzofuranyl	H	37.79	+	–
33	II	C_2H_5	2-Benzofuranyl	H	38.62	+	+

+ Active compounds (compounds having IC_{50} <0.1 μM).
– Inactive compounds.

Table 2. Proposed model based upon superadjacency topochemical index for prediction of anti-HIV activity of HEPT analogs

Model index	Nature of range in proposed model	Index value	Number of compounds falling in the range		Percent accuracy	Average IC_{100} (μM)		Overall accuracy of prediction
			Total	Correct		Total	Correct	
$\int^{Ac}$	Inactive	<37.36	25	23	92.0	12.76	13.86	
	Active	>37.36	8	6	75.0	0.26	0.02	87.87

and has been reported[7] in terms of concentration inhibiting 50% of RT activity (IC_{50} μM). The percent accuracy of prediction was obtained from the ratio of total number of compounds with correctly predicted activity to that of total number of compounds in the active or inactive ranges. The percent classification was obtained from the ratio of the number of compounds in the active and inactive ranges to the total number of compounds present in the dataset (Table 2).

Retrofit analysis of the data presented in the Tables 1 and 2 reveals the following with regard to the superadjacency topochemical index.

- All the 33 compounds in the dataset were classified as either active or inactive (100% classification).
- Out of the 33 compounds, the activity of 29 compounds has been predicted correctly giving 87.87% overall accuracy of prediction.
- The inactive range has superadjacency topochemical index values less than 37.36. 23 out of 25 compounds in the inactive range were predicted correctly. The correctly predicted compounds in this range had average IC_{50} (μM) of 13.86.
- The active range has superadjacency topochemical index values greater than or equal to 37.36. 6 out of 8 compounds in the active range were predicted correctly. The correctly predicted compounds in this range had average IC_{50} (μM) of 0.02.
- For estimation of the IC_{50} (μM) of the new compounds, the following model has been developed.

$$IC_{50}(\mu M) = 0.0133\left(\int^{Ac}\right)^2 - 1.0671\left(\int^{Ac}\right) + 21.4041$$

A topological/topochemical index is a descriptor of molecular structure and is sensitive to such key constitutional features as size, shape, symmetry, branching and heterogeneity of the molecule.[9,10] These computational tools not only provide an alternative to expensive, time consuming and animal sacrificing research but can also helps in faster detection of most favorable compounds.[11] Analysis of the structures of the active compounds reveals that ethyl group as substituent R′ (Basic structure II, Fig. 1.) is a structural feature common to all the active compounds in this dataset, example compound number 24 is inactive while compound number 25 is active. It indicates that this is an important site for interaction with RT. Presence of methyl group at R″ does not seem to offer any advantage. Another benefit offered by this model is that the correctly predicted compounds in the active range have average IC_{50} (μM) of just 0.02 indicating that the active range comprises of very active compounds. Appreciable predictability of the proposed model makes it useful for providing lead structures for the development of potent anti-HIV agents.

References and notes

1. Turker, L. *Indian J. Chem.* **2003**, *42A*, 1442.
2. Trinajstic, N. *Chemical Graph Theory*, 2nd ed.; CRC Press: Boca Raton, FL, 1992.
3. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Study Press: Letchworth, 1986.
4. Estrada, E.; Patlewicz, G.; Uriarte, E. *Indian J. Chem.* **2003**, *42A*, 1315.
5. Ivanciuc, O. In Devillers, J., Balaban, A. T., Eds.; Topological indices and related descriptors in QSAR and QSPR; Gordon & Breach: Amsterdam; 1999; pp 697–777.
6. Bajaj, S.; Sambi, S. S.; Madan, A. K. *QSAR & Comb. Sci.* **2004**, *23*(7), 506.
7. Pontikis, R.; Benhida, R.; Aubertin, A.; Grierson, D.; Monneret, C. *J. Med. Chem.* **1997**, *40*, 1845.
8. Bajaj, S.; Sambi, S. S.; Madan, A. K. *Bioorg. Med. Chem.* **2004**, *12*, 3695.
9. Khadikar, P. V.; Lukovits, I.; Agarwal, V. K.; Shrivastava, S.; Jaiswal, M.; Gutman, I.; Karmarkar, S.; Shrivastava, A. *Indian J. Chem.* **2003**, *42A*, 1436.
10. Bajaj, S.; Sambi, S. S.; Madan, A. K. *J. Mol. Str. (Theochem)*, in press.
11. Grover, M.; Singh, B.; Bakshi, M.; Singh, S. *Pharm. Sci. Technol. Today* **2000**, *3*, 28.