J Incl Phenom Macrocycl Chem (2007) 58:321–326
DOI 10.1007/s10847-006-9160-8

ORIGINAL ARTICLE

Topological models for prediction of adductability of substituted cyclic organic compounds in urea

Seema Thakral · A. K. Madan

Received: 14 July 2006 / Accepted: 31 October 2006 / Published online: 6 March 2007
© Springer Science+Business Media B.V. 2007

Abstract The relationship of urea adductability of substituted cyclic organic compounds with topological descriptors has been investigated. *Wiener's index*—a distance-based topological descriptor, *molecular connectivity index*—an adjacency-based topological descriptor and *eccentric connectivity index*—an adjacency-cum-distance based topological descriptor were employed for the present study. A data set comprising of 45 cyclic organic compounds was utilized. The values of all the three topological indices for every compound involved in the data set were computed using in-house computer program. The resultant data was analyzed and suitable models were developed after identification of adductible ranges. Subsequently, each compound in the data set was classified using these models either as urea adductible or non-adductible, which was then compared with the reported adductability in urea. Accuracy of prediction was found to vary from a minimum of ~90% for a model based upon eccentric connectivity index to a maximum of ~92% for model based upon Wiener's index. Statistical analysis revealed the selected topological indices to be weakly or appreciably intercorrelated for the said data set.

Keywords Eccentric connectivity index · Molecular connectivity index · Wiener's index · Topological descriptors · Urea adductability · Urea inclusion compounds

S. Thakral
GVM College of Pharmacy, Sonipat 131 001, India

A. K. Madan (✉)
Faculty of Pharmaceutical Sciences, M.D. University, Rohtak 124 001, India
e-mail: madan_ak@yahoo.com

Introduction

Among the wide range of solid inclusion compounds that have been studied in recent years, the urea inclusion compounds have received particular attention in view of the wide range of fundamental physicochemical phenomena that they exhibit, including incommensurate structural properties, order–disorder phase transitions, molecular motion and properties relating to one-dimensional confinement [1, 2]. The crystal structure of urea itself has been reported to correspond to tetragonal space group $P42_1m$ [3]. However, in the presence of a suitable guest moiety, urea molecules form an extensively hydrogen-bonded host structure as parallel, non-intersecting, one-dimensional hexagonal tunnels described by space group $P6_122$ (right-handed spiral) or $P6_522$ (left handed spiral) at ambient temperature [4, 5]. The guest molecules are loosely held within these tunnels and are expected to adopt a linear, extended conformation [6] while, the conventional urea inclusion compounds are characterized by hexagonal host tunnel structure, urea is also known to crystallize in modified symmetries in presence of some of the guest moieties e.g., rhombohedral space group *R*32/C with α,ω-disubstituted alkanes [7] or trioxane [8], orthorhombic space group *Pbcn* in presence of 1-4-dichlorobutane [9]. Hence, urea inclusion compounds have been exploited as means of isolating molecules in such conformation and also as prototypical materials for studying various properties of molecules in the particular conformation

[10]. This unique property of selective adduction has led to several industrial applications of urea inclusion compounds, the most important of them being separation of hydrocarbon mixtures in petroleum industry [11, 12].

Urea is known to form inclusion compounds with straight chain compounds having a carbon backbone above a minimum length for a homologous series. In terms of degree of branching or substitution in a guest molecule, little or no branching or substitution is usually a requirement for urea inclusion compound formation. However, guest molecules with non-hydrogen substituents at a position other terminal carbon atoms also form inclusion compounds provided certain steric requirements are fulfilled [13]. As for cyclic compounds, in general, molecules containing benzene or cyclohexane rings do not form inclusion complexes with urea, presumably because these structural components are too wide to fit comfortably inside the tunnel [1]. However 1-phenyloctadecane forms an adduct with urea. The long chain of this compound is readily adducted and apparently the unit cell can easily withstand the distortion caused by an occasional benzene group [14]. Other cyclic groups can be expected to be included in a similar manner by causing occasional distortion of unit cells in a hexagonal structure. Suitable models for prediction of adductability of branched aliphatic compounds in urea were developed earlier [15]. In the present study, suitable models have been developed for prediction of adductability of substituted cyclic organic compounds in urea utilizing topological descriptors as structural parameters.

Topological descriptors are the numerical values associated with chemical constitution and are exploited for correlation of chemical structure with various physical properties, chemical reactivity or biological activity [16]. Graph theory has been applied successfully through translation of chemical structures into characteristic numerical descriptors by resorting to graph invariants. Since the structure of an analogue depends on the connectivity, adjacency and interatomic distances of its constituent atoms, the numerical graph invariants derived from this information can reveal the structural or sub-structural information of a molecule [16]. In recent years a large number of topological indices have been reported and utilized for chemical documentation, isomer discrimination, study of molecular complexity, chirality, similarity/dissimilarity, QSAR/QSPR, drug design and database selection, lead optimization, rational combinatorial library design and for deriving multilinear regression models [16–18]. Topological indices that have been extensively used for structure activity relationship studies include Wiener's index [19, 20], Balaban's indices [21], Hosoya's index [22], Molecular connectivity index [23], Zagreb indices [24] and eccentric connectivity index [25, 26]. In the present study, relationship between topological descriptors of diverse nature and adductability of cyclic compounds in urea has been investigated and suitable models developed for the prediction of adductability of cyclic compounds in urea. The three topological descriptors selected for the present study were molecular connectivity index (χ)—an adjacency based topological index, Wiener's number (W)—a distance-based topological index and eccentric connectivity index (ξ^c)—a adjacency-cum-distance based topological index.

Methodology

Calculation of topological indices

Molecular connectivity index

One of the pioneering topological indices widely applicable to structure-activity/property studies, is molecular connectivity index, which was originally referred to as branching index [23]. It was later expanded to broaden its applications and the chemical literature now uses the term molecular connectivity index [26]. It is denoted by χ and is defined as the sum over all the edges (ij) as per following:

$$\chi = \sum_{i=1}^{n} \sum_{j=1} (V_i V_j)^{-1/2} \quad (1)$$

where V_i and V_j are the degrees of adjacent vertices i and j and n is the number of vertices in a hydrogen depleted molecular structure.

Wiener's index

It is the first reported and widely used topological index in Chemistry [19] and is defined as half sum of the distances between all the pairs of vertices in a hydrogen depleted molecular graph.

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

where P_{ij} is the length of the path that contains the least number of edges between vertex i and vertex j in graph G and n is the maximum possible number of i and j.

Eccentric connectivity index

The eccentric connectivity index [25], an adjacency-cum-distance based topological index denoted by ξ^c, is defined as the summation of the product of eccentricity and the degree of each vertex in the hydrogen depleted molecular graph having n vertices.

$$\xi^c = \sum_{i=1}^{n} (E_i * V_i) \quad (3)$$

where E_i is the eccentricity of the vertex i, V_i is the degree of vertex i and n is the number of the vertices in graph G. The eccentricity E_i of a vertex i in a graph G is the path length from vertex i to vertex j that is farthest from i (E_i = max d (ij); $j\varepsilon$ G); the eccentric connectivity index takes into consideration the eccentricity as well as valency of the vertices in a hydrogen depleted graph.

Model design

A dataset consisting of 45 n cyclic organic compounds was used for the present study [27–33]. The dataset comprised of both the adductible and non-adductible compounds with respect to urea as a host. Values of χ, W and ξ^c of all the cyclic compounds were computed using an in-house computer program. Resulting data was analyzed and suitable models were developed after identification of adductible ranges by maximization of the moving average with respect to the adductible compounds (≤35% = non-adductible, 35–65% = transitional, ≥65% = adductible) in a manner similar to that adopted for identification of active ranges in structure-activity relationship (SAR) studies [34]. Subsequently each compound involved in the data set was classified either as adductibile or non-adductible using these models, which was then compared with the reported adductability in urea [27–33]. The percentage degree of prediction for a particular range was derived from the ratio of the number of compounds classified correctly to the total number of compounds present in that range. The overall degree of prediction was calculated from the ratio of the total number of compounds classified correctly to that of the total number of compounds present in both the adductible and non-adductible ranges.

Intercorrelation between χ, W and ξ^c for all the compounds present in the data set was also studied. The results are summarized in Tables 1 and 2 and Figs. 1–3.

Result and discussions

Graph invariants play a vital role in SPR and SAR studies by quantification of qualitative chemical structures and thereby facilitating obvious correlation with quantified physical/chemical property or biological activity [35]. Graph invariants are numerical values derived from topological representation of hydrogen-suppressed molecular graph and hence characterize specific aspect of a molecule [16]. Though a large number of graph invariants belonging to various categories have been reported in literature, only a handful of them have been successfully employed in structure-activity/property relationships. Recently, suitable models have been developed for prediction of adductability of diverse range of branched aliphatic compounds in urea as the host lattice [15]. In the present study, *Wiener's index W*—a distance-based topological descriptor, *molecular connectivity index* χ—an adjacency-based topological descriptor and *eccentric connectivity index* ξ^c—an adjacency-cum-distance based topological descriptors were successfully utilized for development of models for prediction adductability of substituted cyclic organic compounds in urea.

Retrofit analysis of data in Tables 1 and 2 reveals the following information:

Model based upon χ:

- A total of 33 out of 36 compounds were classified correctly in both adductible and non-adductible ranges. The overall accuracy of prediction was found to be ~92% with regard to adductability in urea.
- The adductible range had a χ value of >9.32. Accuracy of prediction of this range was 91%.
- The non-adductible χ range for cyclic organic compounds was <6.84. Adductability of total of 23 compounds out of 25 (92%) was classified correctly.
- The non-adductible range was ideally separated from the adductible range by a transitional range with χ values varying from 6.84 to 9.32 indicating a gradual transition from non-adductibility to adductibility in urea and vice-versa.

Model based upon W:

- A total of 35 out of 38 cyclic compounds were classified correctly in both adductible and non-adductible ranges. The overall accuracy of prediction was found to be ~92% with regard to adductability in urea.

Table 1 Relationship between topological descriptors and adductability of various cyclic organic compounds in urea

No.	Compound name	X	W	ξ^c	Predicted adductability			Reported adductability	References
					χ	W	ξ^c		
1	Pinene	4.62	122	74	–	–	–	–	[27]
2	Camphene	4.98	123	73	–	–	–	–	[27]
5	Cyclohexane	3.00	27	36	–	–	–	–	[27]
4	Benzene	3.00	27	36	–	–	–	–	[31]
5	1-Phenyl octane	6.93	399	239	±	–	–	–	[28]
6	Cyclopantane	2.50	15	22	–	–	–	–	[27]
7	Methylcyclopentane	2.98	26	29	–	–	–	–	[27]
8	Ethylcyclopentane	3.43	43	44	–	–	–	–	[27]
9	1,1-Dimethyl cyclopentane	3.23	39	34	–	–	–	–	[27]
10	1-Methyl 1-ethyl cyclopentane	3.77	59	49	–	–	–	–	[27]
11	1-Methyl cyclohexane	3.39	42	45	–	–	–	–	[27]
12	1-Ethyl cyclohexane	3.93	64	62	–	–	–	–	[27]
13	1-Cyclopentyl nonane	6.93	414	241	±	±	±	+	[28]
14	1-Phenyl eicosane	12.93	2797	923	+	+	+	+	[28]
15	1-Cyclohexyl octane	6.93	399	239	±	–	–	–	[28]
16	Phenyl octanoate	7.83	563	294	±	±	±	+	[29]
17	Benzyl decanoate	9.33	982	438	+	+	+	+	[29]
18	Benzyl dodecanoate	12.33	1354	552	+	+	+	+	[29]
19	Cyclohexyl dodecanoate	9.83	1161	494	+	+	+	+	[29]
20	Decyl phenyl acetate	9.83	1161	494	+	+	+	+	[29]
21	Phenyl dodecanoate	9.83	1161	494	+	+	+	+	[29]
22	Ethyl benzoate	5.34	164	116	–	–	–	–	[30]
23	*n*-Octadecyl benzoate	13.34	3228	962	+	+	+	–	[29]
24	Phenyl hexanoate	6.83	364	212	–	–	–	–	[29]
25	Cyclohexyl octanoate	7.83	563	294	±	±	±	–	[29]
26	Cyclohexyl decanoate	8.83	826	388	±	±	±	–	[29]
27	Benzyl octanoate	8.33	686	338	±	±	±	–	[29]
28	Methyl salicylate	5.25	152	99	–	–	–	–	[30, 33]
29	1-Phenyl octadecane	11.93	2184	779	+	+	+	+	[14]
30	1-Cyclohexyl eicosane	12.93	2797	923	+	+	+	+	[28]
31	Chlorocyclohexane	3.39	42	45	–	–	–	–	[27]
32	Cyclopentanol	2.89	26	29	–	–	–	–	[27]
33	1-4Dioxane	3.00	27	36	–	–	–	+	[32]
34	Cyclohexanol	3.39	42	45	–	–	–	–	[27]
35	Cyclohexanone	3.39	42	45	–	–	–	–	[27]
36	Borneol	4.98	123	73	–	–	–	–	[27]
37	Cyclopentanone	2.89	26	29	–	–	–	–	[27]
38	Menthane	4.70	122	88	–	–	–	–	[27]
39	α-Lipoic methyl ester	6.33	329	187	–	–	–	+	[33]
40	1,1,2-Trimethylcyclopentane	3.63	56	41	–	–	–	–	[27]
41	Benzyl hexanoate	6.83	364	212	–	–	–	–	[29]
42	1-Phenyl decane	7.93	612	323	±	±	±	–	[13]
43	Naphthyl palmitate	14.81	3772	125.73	+	+	–	+	[29]
44	Quinol dibutyrate	8.65	739	340	+	±	±	+	[29]
45	Quinol dihexanoate	10.65	1389	540	+	+	+	+	[29]

Note: + Adductible compounds, – non-adductible compounds, ± compounds in the transitional range, where adductability could not be specifically assigned

- The adductible range had W values of >981. Accuracy of prediction was of this range was found to be ~91%.
- A non-adductible range with values of <400 was identified. Adductability of a total of 25 compounds out of 27 (~93%) was classified correctly.
- The non-adductible range was ideally separated from the adductible range by a transitional range with W values varying from 400 to 981 indicating a gradual transition from non-adductibility to adductibility in urea and vice-versa.

Model based upon ξ^c:

- A total of 34 out of 38 compounds were classified correctly in both adductible and non-adductible ranges. The overall accuracy of prediction was found to be ~90% with regard to adductability with urea.

Table 2 Topological models for adductability of substituted cyclic organic compounds in urea

Model index	Nature of range in the proposed model	Index value	No. of compounds in the range		Percent accuracy	Overall accuracy
			Total	Correctly predicted		
χ	Non-adductible	<6.84	25	23	92%	92%
	Transitional	6.84–9.32	09	NA	NA	
	Adductible	>9.32	11	10	91%	
W	Non-adductible	<400	27	25	92.60%	92%
	Transitional	400–981	7	NA	NA	
	Adductible	>981	11	10	91%	
ξ^c	Non-adductible	<240	28	25	89%	90%
	Transitional	240–437	7	NA	NA	
	Adductible	>437	10	9	90%	

Note: NA Not applicable

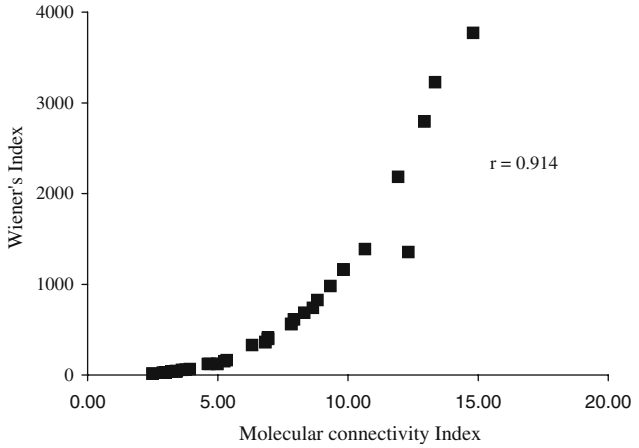

Fig. 1 Scatter plot between the values of Wiener's index and Molecular connectivity index for cyclic organic compounds

- The adductible range had ξ^c values of >437. Accuracy of prediction of this range was found to be 90%.
- The non-adductible ξ^c range was identified as <240. Adductability of total of 25 out of 28 compounds (~89%) was classified correctly.

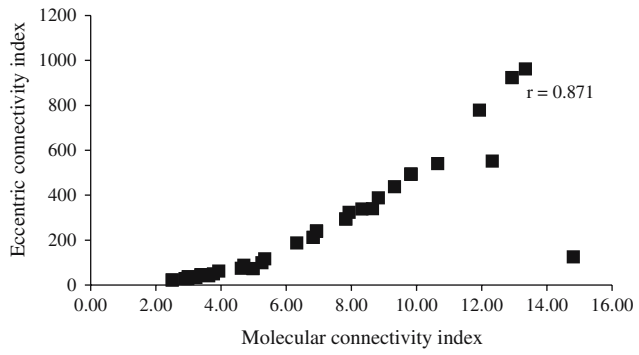

Fig. 2 Scatter plot between the values of Eccentric connectivity index and Molecular connectivity index for cyclic organic compounds

- The non-adductible range was ideally separated from the adductible range by a transitional range indicating a gradual transition from non-adductibility to adductibility in urea and vice-versa.

Statistical analysis in Figs. 1–3 show that indices χ and W are appreciably correlated and pair of indices χ and ξ^c as well as W and ξ^c are weakly correlated ($0.9 \leq r \leq 0.97$ appreciably correlated; $0.50 \leq r \leq 0.89$ weakly correlated) [35].

Excellent correlation between topological descriptors and adductability of substituted cyclic organic compounds in urea has led to development of models for prediction of adductability of substituted cyclic organic compounds in urea. These models with an accuracy of ~ 90–92% can prove be highly beneficial to organic chemists for wide range of industrial applications of diverse nature.

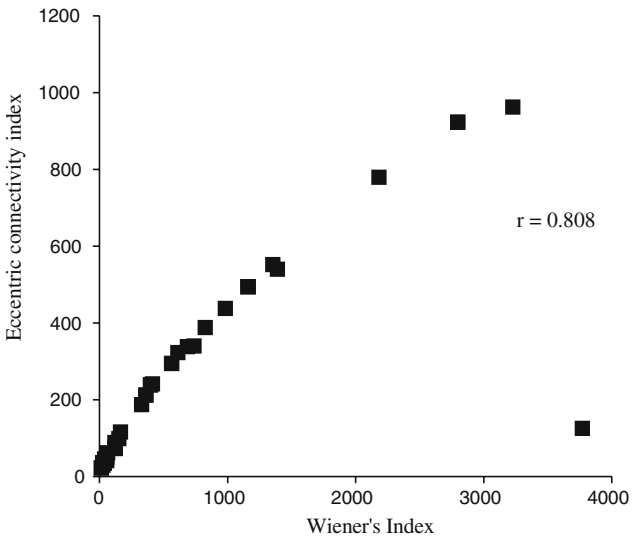

Fig. 3 Scatter plot between the values of Eccentric connectivity index and Wiener's index of cyclic organic compounds

References

1. Hollingsworth, M.D., Harris, K.D.M.: Urea inclusion compounds. In: Atwood, J.L., Davis, J.E.D., Macnicol, D.D., Vogtle, F. (eds.) Comprehensive Supramolecular Chemistry. Solid State Supramolecular Chemistry-Crystal Engineering, vol. 6, pp. 177–237. Pergamon Press, Oxford (1996).
2. Harris, K.D.M.: Urea Inclusion Compounds. In: Atwood, J.L., Steed, J.W. (eds.) Encyclopedia of Supramolecular Chemistry, vol. 2, pp. 1538–1549. Marcel Dekker, New York (2004).
3. Vaughan, P., Donohue, J.: The structure of urea. Interatomic distances and resonance in urea and related compounds. Acta Crystallogr. **5**, 530–535 (1952).
4. Smith, A.E.: The crystal structure of urea–hydrocarbon complexes. Acta Crystallogr. **5**, 224–235 (1952).
5. Harris, K.D.M., Thomas, J.M.: Structure aspects of urea inclusion compounds and their investigations by x-ray diffraction: a general discussion. J. Chem. Soc., Faraday Trans. **86**, 2985–2996 (1990).
6. Frank, S.G.: Inclusion compounds. J. Pharm. Sci. **64**, 1585–1604 (1975).
7. Harris, K.D.M., Smart, S.S., Hollingsworth, M.D.: Structural properties of α, ω-dibromoalkane/urea inclusion compounds: a new type of interchannel guest molecule ordering. J. Chem. Soc. Faraday Trans. **87**, 3423–3429 (1991).
8. Clement, R., Mazieres, C., Guibe, L.: Low temperature phase transition in urea-trioxane inclusion compounds. J. Solid State Chem. **5**, 436–440 (1972).
9. Lenne, H.U., Mez, H.C., Schlenk, W.: The lengths of molecules in inclusion channels of urea and thiourea. Liebigs Ann. Chem. **73**, 70–96 (1968).
10. Smart, S.S., Baghdadi, A.E., Guillaume, F., Harris, K.D.M.: Conformational and vibrational properties of α, ω-dihalogenoalkane/urea inclusion compounds: a Raman scattering investigation. J. Chem. Soc. Faraday Trans. **90**, 1313–1322 (1994).
11. Hayes, D.G., Alstine, J.V., Setterwall, F.: Urea-based fractionation of fatty acids and glycerides of polyunsaturated and hydroxy fatty acid seed oils. J. Am. Oil Chem. Soc. **77**, 207–215 (2000).
12. Bishop, R., Dance, I.G.: New type of helical inclusion networks. Top. Curr. Chem. **149**, 139–188 (1988).
13. Takemoto, K., Sonoda, N.: Inclusion compounds of urea, thiourea and selenourea. In: Atwood, J.W., Davis, J.E.D., MacNicol, D.D. (eds.) Inclusion Compounds, vol. 2, pp. 47–67. Academic Press, London (1984).
14. Findlay, R.A.: Adductive crystallization. In: Schoen, H.M., Mcketta, J.J. (eds.) Interscience Library of Chemical Engineering and Processing. New Chemical Engineering Separation Techniques, vol. 1, pp. 257–318. Interscience Publishers, New York (1962).
15. Thakral, S., Madan, A.K.: Topological models for prediction of adductability of branched aliphatic compounds in urea. J. Inclusion Phemon. Macrocyclic Chem. **56**, 405–412 (2006).
16. Estrada, E., Uriate, E.: Recent advances on the role of topological indices in drug discovery research. Current Med. Chem. **8**, 1573–1588 (2001).
17. Nikolic, S., Kovacevic, G., Milicevic, A., Trinajstic, N.: The Zagreb Indices: thirty years after. Croat. Chem. Acta **76**, 113–124 (2003).
18. Gozalbes, R., Doucet, J., Derouin, F.: Application of topological descriptors in QSAR and drug design: history and new trends. Curr. Drug Targets Infect. Disord. **2**, 93–102 (2002).
19. Wiener, H.: Correlation of heats of isomerization, and differences in heats of vaporization of isomers, among the paraffin hydrocarbons. J. Am. Chem. Soc. **69**, 2636–2638 (1947).
20. Wiener, H.: Structural determination of paraffin boiling points. J. Am. Chem. Soc. **69**, 17–20 (1947).
21. Balaban, A.T.: Topological indices based on topological distances in molecular graphs. Pure Appl. Chem. **55**(2), 199–206 (1983).
22. Hosoya, H.: Topological indices as a sorting device for coding chemical structures. J. Chem. Doc. 12, 181–183 (1972).
23. Randic, M.: On characterization of molecular branching. J. Am. Chem. Soc. **97**(23), 6609–6615 (1975).
24. Gutman, I., Randic, M.: Algebraic characterization of skeletal branching. Chem. Phys. Lett. **47**, 15–19 (1977).
25. Sharma, V., Goswami, R., Madan, A.K.: Eccentric connectivity index: a novel highly discriminating topological descriptor for structure-property and structure-activity studies. J. Chem. Inf. Comput. Sci. **37**, 273–282 (1997).
26. Kumar, V., Madan, A.K.: Topological models for the prediction of cyclindependent kinase 2 inhibitory activity of aminothiazoles. MATCH Commun. Math. Comput. Chem. **51**, 59–78 (2004).
27. Swern, D.: Urea and thiourea complexes in separating organic compounds. Ind. Eng. Chem. **47**, 216–221 (1955).
28. Schiessler, R.W., Flitter, D.: Urea and thiourea adduction of C_5–C_{42}-hydrocarbons. J. Am. Chem. Soc. **74**, 1720–1723 (1950).
29. Truter, E.V.: Urea complexes of some branched-chain and cyclic esters. J. Chem. Soc. 2416–2419 (1951).
30. Linstead, R.P., Whallwy, M.: The formation of crystalline complexes between urea and esters, and their application to separation of mixtures of esters. J. Chem. Soc. 2987–2989 (1950).
31. Zimmerschied, W.J., Dinerstein, R.A., Wietkamp, A.W., Marschner, R.F.: Complexes of urea with linear aliphatic compounds. J. Am. Chem. Soc. **71**, 2947 (1949).
32. Redlich, O., Gable, C.M., Dunlop, A.K., Millar, R.W.: Addition compounds of urea and organic substances. J. Am. Chem. Soc. **72**, 4153–4160 (1950).
33. Mima, H., Nishikawa, M.: Inclusion compounds of α-lipoic acid methyl ester with urea and thiourea. J. Pharm. Sci. **53**, 931–934 (1964).
34. Gupta, S., Singh, M., Madan, A.K.: Predicting anti-HIV activity: computational approach using a novel topological descriptor. J. Comput. Aided Mol. Des. **15**, 671–678 (2001).
35. Gupta, S., Singh, M., Madan, A.K.: Application of graph theory: relationship of molecular connectivity index and atomic molecular connectivity index with anti-HSV activity. J. Mol. Struct. (Theochem) **571**, 142–152 (2000).