

*Chapter-8*

*Development of*

*Mathematical Models for*

*Prediction of Heat of*

*Decomposition of Urea*

*Inclusion Compounds*

---

---

## **8.1 Introduction**

Urea, in the presence of a suitable guest moiety, crystallizes into an approximately hexagonal framework with long, open, essentially infinite, parallel and non-intersecting tunnels or channels of diameter of 5.5 – 5.6Å, which completely enclose guest molecules (Harris and Thomas, 1990). Urea belongs to the category of soft hosts as the hexagonal host structure undergoes substantial structural reorganization when the guest component is removed (Harris, 1997). Linear guest molecules are included along these channels in an extended planar zigzag conformation with their long axis parallel to the *c* axis. Structure compatibility between the host and guest components is fundamental to inclusion phenomena and consequently urea is known to form inclusion compounds with guest molecules that are based on a sufficiently long alkane with only a limited degree of substitution allowed (Takemoto and Sonoda, 1984). In general, highly substituted molecules or molecules containing cyclic moieties do not form inclusion compounds with urea. Thus the important criteria, which determine whether a particular guest structure would be included in urea tunnels, is the degree of branching or substitution of the guest moiety. However, slightly branched molecules are included in urea but substituents, which are not part of the linear chain, distort the host lattice (Findlay, 1962; Fetterly, 1964) and significant changes are induced within the urea channels to accommodate a substituted guest moiety owing to the soft nature of urea host (Harris, 1997). On the other hand, the formation of urea inclusion compounds is exothermic in nature, which clearly indicates that the resulting compounds are stable. However, substituents, which do not form part of the linear chain, will naturally lead to distortion and weakening of host structure comprising of narrow channels with consequent decrease in heat of decomposition of urea inclusion compounds.

In the present study, QSPR models are developed for quantitative estimation of the heat of decomposition of a urea inclusion compound on the basis of the structural features of the guest molecule utilizing molecular topological descriptors.

The basic rationale behind use of QSPR is the fact within particular family of chemical compounds, especially of organic chemistry; there are strong correlations between structure and observed property (Katrinzy *et al*, 1995). The inherent problem in development of a QSPR can be attributed to non-quantitative nature of chemical

structures. This problem can be easily overcome by translation of molecular structures into numerical values, by use of the topological descriptors (Sabljic and Trinajstic, 1981). Hence the principal objective is to find a way to represent molecules as simple numbers and to reduce the problem of structure-property relationship to a problem of correlation between two series of numbers, one of them representing the property and the other representing the chemical structure (Hansen and Jurs, 1888). Molecular descriptors calculated from topological representations are called topological descriptors (Estrada and Uriate, 2001). These descriptors are expected to contain important information on molecules that can be useful in describing physical, chemical and biological properties. Topological descriptors have been extensively used for structure-property relationship studies and it has been shown that for physicochemical properties, topological descriptors seem to be more advantageous than non-topological ones for linear regression analysis (Katritzky and Gordeeva, 1993).

Reasonably good results were obtained for prediction of adductability in urea for a dataset of diverse range of branched aliphatic compounds using three topological descriptors, *molecular connectivity index*- an adjacency based topological descriptor, *Wiener's index*- a distance based topological descriptor, and *eccentric connectivity index*- an adjacency cum distance based topological descriptor (Thakral and Madan, 2006). The relationship of these three descriptors with heat of decomposition of urea inclusion compounds having aliphatic compounds as the guest have been further investigated.

## **8.2 Methodology**

### **8.2.1 Calculations of topological indices**

*Molecular connectivity index* ( $\chi$ ) is defined as the sum over all the edges ( $ij$ ) as per following:

$$\chi = \sum (\delta_i \delta_j)^{-1/2}$$

Where  $\delta_i$  and  $\delta_j$  are the degrees of adjacent vertices  $i$  and  $j$  (Randic, 1975). The vertex degree or valency of atom  $i$  ( $\delta_i$ ) is the number of bonds incident on the vertex  $i$ .

*Weiner's index* ( $W$ ) is defined as sum of the distances between all the pairs of vertices in a hydrogen-suppressed molecular graph

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

where  $P_{ij}$  is the length of the path that contains the least number of edges between vertex  $i$  and vertex  $j$  in the graph  $G$  and  $n$  is the number of vertices in a hydrogen depleted molecular structure (Wiener, 1947).

*Eccentric connectivity index* ( $\xi^c$ ) 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.

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

Where  $E_i$  is the eccentricity of the vertex  $i$ ,  $\delta_i$  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  $i$  ( $E_i = \max_{j \in G} d_{ij}$ );  $d_{ij}$  is the length of the shortest path between vertices  $i$  and  $j$ . The eccentric connectivity index takes into consideration the eccentricity as well as valency of the vertices in a hydrogen depleted graph (Sharma *et al*, 1997).

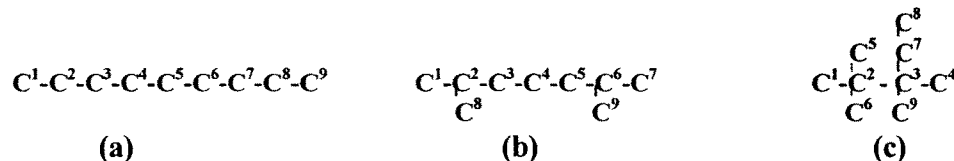
The calculation of molecular connectivity index ( $\chi$ ), Wiener's index ( $W$ ) and eccentric connectivity index ( $\xi^c$ ) for three isomers of nonane has been exemplified in **Fig. 8.1**.

### 8.2.2 Model design

A dataset comprising of 44 aliphatic compounds with the reported heat of decomposition (kcal./mole) of their urea inclusion complex was extracted from the literature (McAdie, 1963; Zimmerschied *et al*, 1950; Redlich *et al*, 1950). The dataset included diverse range of aliphatic compounds *i.e.* alkanes, alkenes, haloalkanes, alcohols, aldehydes, ethers, acids and esters. (**Table 8.1**)

Values of  $\chi$ ,  $W$  and  $\xi^c$  for all the compounds in the dataset were computed using in-house computer program. Subsequently, mathematical models were developed for prediction of heat of decomposition of urea inclusion compounds using these descriptors. The quality of models was expressed by fitted (descriptive) statistical parameters: the correlation

Fig. 8.1 Calculation of values of molecular connectivity index, wiener index and eccentric connectivity index for three isomers of nonane using hydrogen suppressed molecular structure.



Distance matrices (D)

	1	2	3	4	5	6	7	8	9	$E_i$	$S_i$		1	2	3	4	5	6	7	8	9	$E_i$	$S_i$		1	2	3	4	5	6	7	8	9	$E_i$	$S_i$
1	0	1	2	3	4	5	6	7	8	8	36	1	0	1	2	3	4	5	6	2	6	6	29	1	0	1	2	3	2	2	3	4	3	4	20
2	1	0	1	2	3	4	5	6	7	7	29	2	1	0	1	2	3	4	5	1	5	5	22	2	1	0	1	2	1	1	2	3	2	3	11
3	2	1	0	1	2	3	4	5	6	6	24	3	2	1	0	1	2	3	4	2	4	4	19	3	2	1	0	1	2	2	1	2	1	2	12
4	3	2	1	0	1	2	3	4	5	5	21	4	3	2	1	0	1	2	3	3	3	3	18	4	3	2	1	0	3	3	2	3	2	3	19
5	4	3	2	1	0	1	2	3	4	4	20	5	4	3	2	1	0	1	2	4	2	4	19	5	2	1	2	3	0	2	3	4	3	4	20
6	5	4	3	2	1	0	1	2	3	5	21	6	5	4	3	2	1	0	1	5	1	5	22	6	2	1	2	3	2	0	3	4	3	4	20
7	6	5	4	3	2	1	0	1	2	6	24	7	6	5	4	3	2	1	0	6	2	6	29	7	3	2	1	2	3	3	0	1	2	3	17
8	7	6	5	4	3	2	1	0	1	7	29	8	2	1	2	3	4	5	6	0	6	6	29	8	4	3	2	3	4	4	1	0	3	4	24
9	8	7	6	5	4	3	2	1	0	8	36	9	6	5	4	3	2	1	2	6	0	6	29	9	3	2	1	2	3	3	2	3	0	3	19

Adjacency Matrices (A)

	1	2	3	4	5	6	7	8	9	$\delta_i$		1	2	3	4	5	6	7	8	9	$\delta_i$		1	2	3	4	5	6	7	8	9	$\delta_i$	
1	0	1	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	1
2	1	0	1	0	0	0	0	0	0	2	2	1	0	1	0	0	0	0	1	0	0	3	2	1	0	1	0	1	1	0	0	0	4
3	0	1	0	1	0	0	0	0	0	2	3	0	1	0	1	0	0	0	0	0	0	2	3	0	1	0	1	0	0	1	0	1	4
4	0	0	1	0	1	0	0	0	0	2	4	0	0	1	0	1	0	0	0	0	0	2	4	0	0	1	0	0	0	0	0	0	1
5	0	0	0	1	0	1	0	0	0	2	5	0	0	0	1	0	1	0	0	0	0	2	5	0	1	0	0	0	0	0	0	0	1
6	0	0	0	0	1	0	1	0	0	2	6	0	0	0	0	1	0	1	0	1	3	6	0	1	0	0	0	0	0	0	0	1	
7	0	0	0	0	0	1	0	1	0	2	7	0	0	0	0	0	1	0	0	0	1	3	7	0	0	1	0	0	0	0	1	0	2
8	0	0	0	0	0	0	1	0	1	2	8	0	1	0	0	0	0	0	0	0	1	3	8	0	0	0	0	0	0	1	0	0	1
9	0	0	0	0	0	0	0	1	0	1	9	0	0	0	0	0	1	0	0	0	1	3	9	0	0	1	0	0	0	0	0	0	1

Molecular connectivity index

$$\chi = \sum (\delta_i \delta_j)^{1/2} = 4.414 \qquad = 4.125 \qquad = 3.810$$

Wiener Index

$$W = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n P_{ij} = 120 \qquad = 108 \qquad = 81$$

Eccentric Connectivity Index

$$\xi^c = \sum_{i=1}^n (E_i * \delta_i) = 96 \qquad = 76 \qquad = 48$$

coefficient  $r_{\text{fit}}$ , the standard error of estimate  $s_{\text{fit}}$  and  $F$  is the result of Fischer's test. The models were internally cross-validated by a leave-one-out procedure (Nikolic *et al*, 2003). Statistical parameters for the cross-validated models were denoted as  $r_{\text{cv}}$  and  $s_{\text{cv}}$  where  $\text{cv}$  stands for cross-validation procedure.

The following mathematical model was developed for the estimation of  $\Delta H$  using molecular connectivity index ( $\chi$ ):

$$\Delta H = 2.652 (\pm 0.155) \chi - 1.969 (\pm 1.117) \quad (\text{Model I})$$

$$N = 44 \quad r_{\text{fit}} = 0.935 \quad s_{\text{fit}} = 2.2979 \quad r_{\text{cv}} = 0.929 \quad s_{\text{cv}} = 2.4053 \quad F = 293.65$$

The following mathematical model was developed for the estimation of  $\Delta H$  using Wiener index ( $W$ ):

$$\Delta H = 0.01185 (\pm 0.001) W + 9.267 (\pm 0.724) \quad (\text{Model II})$$

$$N = 44 \quad r_{\text{fit}} = 0.886 \quad s_{\text{fit}} = 3.0143 \quad r_{\text{cv}} = 0.850 \quad s_{\text{cv}} = 3.4177 \quad F = 153.05$$

The following mathematical model was developed for the estimation of  $\Delta H$  using eccentric connectivity index ( $\xi^e$ ):

$$\Delta H = 0.03667 (\pm 0.002) \xi^e + 6.887 (\pm 0.757) \quad (\text{Model III})$$

$$N = 44 \quad r_{\text{fit}} = 0.917 \quad s_{\text{fit}} = 2.6386 \quad r_{\text{cv}} = 0.904 \quad s_{\text{cv}} = 2.7756 \quad F = 227.15$$

The results indicating index values and predicted  $\Delta H$  for models I-III are compiled in **Table 8.1**.

### **8.3 Result and Discussions**

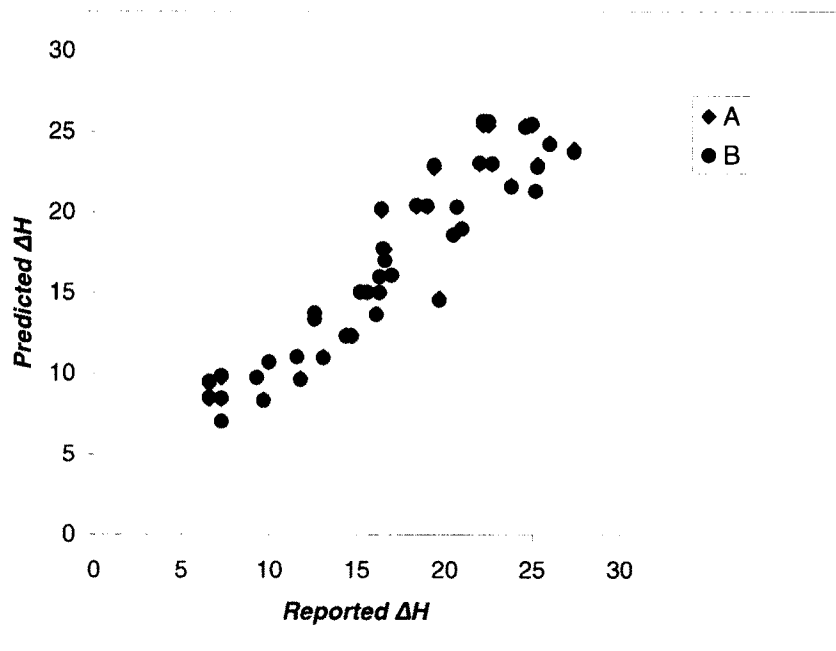
In the present study a set of three commonly used topological descriptors of diverse nature has been utilized for the development of models for the prediction of heat of decomposition of urea inclusion compounds. Retrofit analysis of the mathematical models developed as above indicate that it is possible to establish a reasonably accurate quantitative structure-property relationship to estimate  $\Delta H$  of urea inclusion compound based on a simple set of molecular topological descriptors. As per the value of correlation coefficient  $r$  for these models,  $\Delta H$  of urea inclusion compounds can be considered to be appreciably correlated with the two of the topological descriptors selected for the study i.e. molecular connectivity index ( $\chi$ ) and eccentric connectivity index ( $\xi^e$ ) ( $0.90 \leq r \leq 0.97$  appreciably correlated (Nikolic *et al*, 2003)) while weakly correlated with descriptor  $W$  ( $0.50 \leq r \leq 0.89$  weakly correlated (Nikolic *et al*, 2003)). Judging by the statistical

**Table 8.1** Heat of decomposition  $\Delta H$  of urea inclusion compounds using topological models.

No.	Compound	$\chi$	W	$\xi^c$	Predicted $\Delta H$ using Model			Reported $\Delta H$
					I	II	III	
1	Di-n-butyl ether	4.41	120	96	9.71	10.50	9.99	9.3
2	N-octane	3.91	84	74	8.38	10.07	9.24	9.7
3	N-heptane	3.41	56	54	7.05	9.73	8.56	7.3
4	1-octene	3.91	84	74	8.38	10.07	9.24	7.3
5	Octanol	4.41	120	96	9.71	10.50	9.99	7.3
6	Caprylic acid	4.77	158	111	10.67	10.96	10.50	10
7	1-decene	4.91	165	122	11.04	11.04	10.88	11.6
8	N-nonane	4.41	120	96	9.71	10.50	9.99	11.8
9	Methyl ethyl ketone	2.27	18	19	4.02	9.28	7.37	4.7
10	2-octene	3.91	84	74	8.38	10.07	9.24	6.6
11	Octanol-2	4.27	114	87	9.34	10.43	9.69	6.6
12	Dodecene-1	5.91	286	182	13.70	12.49	12.92	12.6
13	Capric acid	5.77	277	169	13.33	12.39	12.48	12.6
14	N-decane	4.91	165	122	11.04	11.04	10.88	13.1
15	1-undecene	5.41	220	150	12.37	11.70	11.83	14.4
16	1-bromodecane	5.41	220	150	12.37	11.70	11.83	14.7
17	Lauryl bromide	6.41	364	216	15.03	13.43	14.08	15.2
18	Lauryl alcohol	6.41	364	216	15.03	13.43	14.08	15.6
19	Dodecane	5.91	286	182	13.70	12.49	12.92	16.1
20	Adipic acid	4.63	151	122	10.30	10.87	10.88	3.3
21	Lauric acid	6.77	444	239	15.99	14.39	14.87	16.3
22	Lauryl chloride	6.41	364	216	15.03	13.43	14.08	16.3
23	1-bromohexadecane	8.41	816	384	20.35	18.85	19.81	20.7
24	1-bromotetradecane	7.414	560	294	17.70	15.78	16.74	16.5
25	Aminotetradecane	7.414	560	294	17.70	15.78	16.74	16.6
26	Methyl undecylenate	6.81	435	237	16.10	14.28	14.80	17
27	Hexadecanol	8.41	816	384	20.35	18.85	19.81	18.4
28	Aminohexadecane	8.41	816	384	20.35	18.85	19.81	19
29	Octadecanol	9.41	1140	486	23.01	22.74	23.29	22
30	Undecylenic acid	6.27	354	203	14.66	13.31	13.64	19.7
31	Myristic acid	7.77	667	321	18.65	17.07	17.66	20.5

No.	Compound	$\chi$	W	$\xi^c$	Predicted $\Delta H$ using Model			Reported $\Delta H$
					I	II	III	
32	Methyl tetradecanoate	8.31	790	365	20.09	18.54	19.16	16.4
33	Hexadecane	7.91	680	338	19.02	17.22	18.24	21
34	Methyl hexadecanoate	9.31	1110	465	22.75	22.38	22.57	19.4
35	Methyl stearate	10.31	1506	577	25.41	27.13	26.39	22.2
36	Methyl linolenate	10.31	1506	577	25.41	27.13	26.39	22.5
37	1- aminooctadecane	9.41	1140	486	23.01	22.74	23.29	22.7
38	Octadecene	8.91	969	434	21.68	20.69	21.52	23.8
39	Methyl linoleate	10.31	1506	577	25.41	27.13	26.39	24.6
40	Methyl oleate	10.31	1506	577	25.41	27.13	26.39	25
41	Palmitic acid	8.77	954	415	21.31	20.51	20.87	25.2
42	Bromooctadecane	9.41	1140	486	23.01	22.74	23.29	25.3
43	Didecane	9.914	1330	542	24.35	25.02	25.20	26
44	Oleic acid	9.77	1313	521	23.97	24.82	24.48	27.4

Fig. 8.2 Scatter plots between the reported and predicted heat of decomposition of urea inclusion compounds for the fit A and cross-validated B Model I.





parameters, Model I appears to be the best model, which explains almost 87.5% of the variance in the heat of decomposition for the series of compounds included in the study.

Quite proximal values of  $r_{\text{fit}}$  and  $r_{\text{cv}}$  and values of  $s_{\text{fit}}$  and  $s_{\text{cv}}$  indicate the stability of the models. A plot of reported vs predicted  $\Delta H$  for the fit and cross-validated values for Model I is shown in **Fig. 8.2**.

*There is definite relationship between topological descriptors and heat of decomposition of urea inclusion compounds.* The importance of topology of molecules on their supramolecular chemistries and, consequently, on the properties of their aggregates is absolute. Chemical topology, in general, is the connectivity and linkages of atoms and molecules in space. This connectivity and linkages play vital role in determining the chemical and physical characteristics of the molecules (such as boiling point of differently branched alkanes) and assemblies of them (such as inclusion properties of crystals) (Amabilino, 2004).

Topology of molecules not only reveals adjacency but also relative distances between various atoms comprising the molecule. **Figure 8.1** depicts distance matrices (D) and adjacency matrices (A) of three isomers of nonane ranging from purely linear *n*-nonane (Fig. 1(a)) to highly branched isomer of nonane (Fig. 1(c)). As observed from **Fig. 8.1**, the values in both adjacency and distance matrices change significantly with gradual increase in branching, leading to corresponding decrease in the values of molecular connectivity index- an adjacency based topological descriptor, Wiener's index- a distance based topological descriptor, and eccentric connectivity index- an adjacency cum distance based topological descriptor. Urea, which is known to form stable inclusion compounds with long linear compounds, also forms inclusion compounds with slightly branched molecules. Since formation of inclusion compound is dependent upon linearity of guest component, therefore any increase in branching or decrease in linearity will reduce the propensity of adduct formation. *Since the formation of urea inclusion compounds is exothermic in nature, therefore, their stability is naturally related to heat of decomposition.* Thus, branching of guest molecules leads to not only decrease in the possibility of adduct formation, but also to distortion of host lattice in the event of formation of an adduct. *Any distortion of host lattice will weaken the host structure with a corresponding decrease in the heat of decomposition.* Since topological descriptors

### *Prediction of Heat of Decomposition*

encode information regarding adjacency, connectedness, compactness, branching or linearity of a molecule, therefore, these are directly related not only to the propensity of adduct formation but also to the heat of decomposition of urea inclusion compounds.

Thus, QSPR treatment afforded significant topological parameters correlation. Given the molecular structure of a compound, the heat of decomposition of its urea inclusion compound can be easily estimated with the help of these mathematical models. The results are good enough and errors are nearly same as experimental uncertainties.