CHAPTER –2

REVIEW OF LITERATURE
Design, development and commercialization of a drug is a tedious, time consuming and cost-intensive process. Considering both the potential benefits to human health and the enormous cost in time and money of drug discovery, any tool or technique that increases the efficiency of any stage of drug discovery enterprise will be highly prized. Recent estimates have placed the overall cost of successfully developing a single drug entity, from discovery to marketplace, at over $300 million. The potential for this cost to increase even further, considered against the backdrop of cost containment in the health care industry in general and pharmaceuticals in particular has provided incentive to reexamine the drug development process. Structure activity relations can play a vital role in making the drug development process more efficient and cost effective.

Even a cursory review of the SAR/SPR literature indicates the importance of the manner in which chemical structure is represented. The problem in the development of a suitable correlation between chemical structures and properties can be attributed to the non-quantitative nature of chemical structures. Graph theoretic techniques constitute vital mathematical tools for overcoming this problem. Graph theory when applied to SAR essentially involves translation of chemical structures into numerical values in the form of a characteristic polynomial, a matrix, a sequence or a graph invariant. Some general features of graph invariants are that:

1. For computation of graph invariants, hydrogen depleted structure of a molecule is employed.
2. Graph invariants do not depend on arbitrary numbering of the atoms and hence are same for two isomorphic graphs.

3. Stereochemical features of the molecule are disregarded in computation of graph invariants.

4. All graph invariants exhibit varying degree of degeneracy.

There are distinct trends in graph invariant research:

1. The *simultaneous use of more than one graph invariant*, i.e. a superindex better characterizes chemical structure as compared to a single numerical invariant and

2. The *development of invariants* with high discriminating power.

Success of the latter approach is evident from the increasing popularity of different graph invariants in structure-property relationship and structure-activity relationship studies. This is mainly due to the fact that graph descriptors can be calculated for any real or hypothetical graph structure whereas physicochemical parameters used in SPR or SAR are not uniformly available. It has also been pointed out that the non-uniqueness of graph invariants is not a very serious handicap for SPR or SAR.

The figure 2.1 represents an empirical property as a function of ‘a:C \( \rightarrow \) R’ which maps the set ‘C’ of molecules into real line ‘R’. A non-empirical SAR is a composition of description function ‘b\(_1\): C \( \rightarrow \) D’ mapping each chemical structure of ‘C’ into space of non-empirical structural descriptors ‘D’ and a prediction function ‘b\(_2\): D \( \rightarrow \) R’ which maps the descriptors into real line ‘R’. When \([a(C) - b\(_1\)b\(_2\)(C)]\) is within the range of experimental error a good non-empirical predictive model is obtained. PAR is the composition of ‘o\(_1\): C \( \rightarrow \) M’ that maps the set ‘C’ into graph property space ‘M’ and ‘o\(_2\): M \( \rightarrow \) R’ mapping graph properties into the real line."
Figure 2.1: Composition functions for structure-activity relationship and property-activity relationship. (Where $a =$ empirical structure activity relationship, $b_1, b_2 =$ non-empirical structure activity relationship; $o_1, o_2 =$ property activity relationship)
Although various graph invariants have been proposed, only a few of them have been widely used in SAR to take into account steric effects. Various invariants having relevance with SAR/QSPR have been briefly but critically reviewed in the present chapter.

**Adjacency based graph Invariants**

The first class of invariants is based on the consideration that the whole set of connections between adjacent pairs of atoms may be represented in a matrix form, termed as adjacency matrix. The entries $a_{ij}$ of the matrix equals one if vertices i and j are adjacent and zero otherwise.

Platt$^{30}$, in 1947, was the first to use a graph invariant in correlations with certain graph properties. The Platt index ($F$) is defined by

$$F(G) = \sum_{j=1}^{A} e_j$$

Where $e_j$ is the degree of the edge $j$ i.e. the number of edges adjacent to edge $j$.

The Gordon-Scantlebury index ($N_2$) is defined as the number of distinct ways in which the chain fragment $C_3$, called links can be embedded onto the carbon skeleton of a given molecule. This index is related to the $F$ value

$$N_2 = \frac{1}{2} F(G)$$

$$N_2 = \sum P_2$$

Where $P_2$ is the path of length 2.

Gutman and Trinajstic$^{31}$, in 1972, were among the first to be interested in the quantitative characterization of the degree of branching. This structural feature was accounted by vertex degree ($\delta_i$) that is equal to equal to the sum of the entries in
row $i$ in the adjacency matrix. With this assumption two graph invariants $M_1$ and $M_2$

Zagreb group parameters were derived for total $n$ energy

$$M_1 = \sum_{i=1}^{n} \delta_i^2$$

$$M_2 = \sum_{(i,j)} \delta_i \delta_j$$

Where $\delta_i$ is the degree of vertex $i$ and $\delta_i \delta_j$ is the weight of the edge $ij$.

Zagreb group parameter $M_1$ and Gordon-Scantlebury index $N_2$ belong to the same family called quadratic invariants.

Lovasz and Pelikan, in 1973, suggested that the largest (first) eigen value of the characteristic polynomial, matrix of the molecular graph is an invariant. It has been suggested that the leading eigen value $\lambda_1$ of the adjacency matrix of molecular graphs of alkanes represents an index of molecular branching.

Randic, in 1975, introduced a branching index $M_R$ similar to $M_2$ index. Each vertex was assigned a value of $1/(\delta^{1/2})$. The edge index was given a value of the product of the associated vertex values.

$$M_R = \sum_{(i,j)} \frac{1}{\sqrt{\delta_i \times \delta_j}}$$

The branching index was reported to be in agreement with the empirical index of Kovats based on retention data.

Kier and Hall, in 1976, recognized in the Randic procedure the possibility for the development of a more general method to describe organic molecular structures in terms of graph invariants. Assuming $\chi^R = 1^\chi$, where the prefix 1 indicates that the $1^\chi$ index is for one-edge dissection of the structural formula, Kier
and Hall adopted a general scheme to calculate higher-order dissection of the molecular skeleton.

$$\chi = \sum \frac{1}{\sqrt{\delta_i \times \delta_j \times \cdots \times \delta_{n+1}}}$$

Where $\delta_i$ correspond to the coordination number (degree) of vertices.

Molecular connectivity index has following advantages:

1. It possesses great discriminating power by virtue of its high monotonicity.
2. Computation of molecular connectivity index is simple, as only basic algorithm needs to be applied.
3. Molecular connectivity index being related to degree of branching is a good measure of molecular surface area or volume.
4. Molecular connectivity index has provision to consider heteroatoms and multiple bonds.
5. Molecular connectivity index can be applied to acyclic, cyclic and aromatic molecules.
6. Molecular connectivity index correlates well with physicochemical and biological properties.

Kier and Hall\textsuperscript{35}, in 1976, extended the molecular connectivity index to include molecules containing heteroatoms. The $\delta$ value of the heteroatom was modified to take account of its number of attached hydrogen atoms. Therefore the vertex degree ($\delta_i$) is replaced by the vertex connectivity ($\delta_i'$)

$$\delta_i' = Z'' - h_i$$

Where $Z''$ and $h_i$ are respectively, the number of valence electrons and the number of hydrogen atoms bonded to vertex $i$. 
The heteroatoms (O, N, F) connectivity calculation takes into account the adjacent bonded atoms plus all \( n \) and lone pair electrons.

**Kier and Hall**\(^{36}\), in 1978, explored correlation between three molecular connectivity invariants and muscarinic receptor affinity of acetylcholine antagonists. An equation relating these invariants to muscarinic affinity of antagonists was derived and found to be capable of predicting the affinity of other antagonists as well as agonist molecules.

**Burden**\(^{37}\), in 1989, presented a method for generating molecular identification numbers of hydrogen-depleted structures from the eigen values of a connectivity matrix. The use of the method was proposed for unique numbering of the atoms in a structure and for identifying the atoms that belong to different substructures of disconnected main structure.

**Roca and Maria**\(^{38}\), in 1992, applied molecular connectivity index for prediction and interpretation of several pharmacological and microbiological properties of a group of antiviral drugs.

Despite numerous applications molecular connectivity index is susceptible to following limitations:

1. Despite good discriminating power it is not unique e.g. valence values for both ether and ester oxygen are same.
2. Stereoisomers exhibiting different biological activities have same index values.

**Balaban**\(^{39}\), in 1992, classified the invariants based on local vertex invariants (LOVI) in three generations. The first-generation invariants are bookkeeping numbers, with high degeneracy, obtained from integer graph invariants. The second-generation invariants are structural descriptors, with medium to low degeneracy,
obtained from manipulation of local graph properties that are integer or rational numbers. Finally, structural and bookkeeping numbers obtained from graph invariants involving real numbers compose the third-generation of invariants and has low degeneracy.

Estrada\textsuperscript{40}, in 1995, introduced the edge-analog of the classical Randic vertex connectivity index of molecular graph. A square and symmetric edge-adjacency matrix represented adjacency between edges in molecular graphs. The edge connectivity index, a second-generation index, correlated well with molar volume. The index was claimed to be capable of discriminating between isomers.

Estrada\textsuperscript{41}, in 1995, used different weights in the non-diagonal entries of the edge-adjacency matrix as a way to differentiate heteroatoms in the edge connectivity index.

Estrada and Ramirez\textsuperscript{42}, in 1996, defined a topographic edge connectivity index based on edge adjacency relationships. The elements of edge set were substituted by bond orders between connected atoms in the molecule. The proposed index discriminated the isomers, including heteroatoms differentiation and conformational isomerism.

Chang and Xu\textsuperscript{43}, in 1996, proposed a highly discriminating molecular index, EAID, based on the extended adjacency matrix. Its use was recommended as supplementary reference for CAS registry numbers for structure documentation.

Grassy \textit{et al.}\textsuperscript{44}, in 1998, illustrated a new avenue for using graph invariants in rational design of biologically active molecules. They pointed out that

1. No alternative computational approaches are today practical for combinatorial libraries having in excess of 100,000 molecules.
2. Graph invariants along with other molecular descriptors apparently are highly efficient (in filtering two dozen molecules out of over quarter of million).

3. The resulting active molecule was two orders of magnitude more potent than the lead molecule.

Estrada and Rodriguez\textsuperscript{45}, in 1999, demonstrated the independence of edge connectivity index to 40 invariants belonging to first-, second-, and third generation. They concluded that edge connectivity index contains structural information not accounted for by other invariants and can be important for QSPR.

Nikolic \textit{et al.}\textsuperscript{46}, in 2000, symmetry-modified Zagreb invariants $M_1$ and $M_2$ by summing up only degrees (SMM$_1$) or edge weights (SMM$_2$) of symmetry nonequivalent vertices or edges of graphs. On comparing closely related symmetry-independent and symmetry-dependent complexity invariants they produced different ordering.

Basak \textit{et al.}\textsuperscript{47}, in 2000, considered molecular surface dependent properties (boiling point and gas chromatograph retention times) and molecular volume dependent properties (molar volume and molar refraction). They found that edge connectivity invariants were appropriate for structure-molecular volume properties modeling and the vertex connectivity invariants in structure-molecular surface properties.

Gupta \textit{et al.}\textsuperscript{48}, in 2000, conceptualized a novel topological descriptor termed as connective eccentricity index and investigated its discriminating power with regard to antihypertensive activity of N-benzylimidazole derivatives.

Bonchev\textsuperscript{49}, in 2001, predicted physicochemical properties of alkane compounds with molecular connectivity and two probe connectivity functions – the
overall connectivity versions of the second Zagreb group index, and a derivative inverse function of this index. The favorable comparisons indicate potential of the overall connectivity indices for QSAR/QSPR applications.

Lukovits and Linert\textsuperscript{50}, in 2001, applied a chiral function $F$ satisfying the condition $F(D) = -F(L)$, where $D$ and $L$, denote enantiomers of the same structure, in combination with Randic atom-modified first order molecular connectivity index.

Mannhold and van de Waterbeemd\textsuperscript{51}, in 2001, used whole molecule approaches to inspect the entire molecule by applying molecular lipophilicity potentials (MLP), topological indices or molecular properties to quantify log P.

Gupta \textit{et al.}\textsuperscript{52}, in 2001, conceptualized a novel topological descriptor termed as eccentric adjacency index for predicting biological activity. Eccentric adjacency index was successfully employed for development of model for prediction of anti-HIV activity of derivatives of 1-[(2-hydroxyethoxy) methyl]-6-(phenylthio) thymine (HEPT).

Kezele \textit{et al.}\textsuperscript{53}, in 2001, tested the use of variable connectivity index in QSAR. The index yielded very good regression equations in the case of homogenous sets of molecules.

Quigley and Lloyd\textsuperscript{54}, in 2002, correlated various physicochemical properties with the first order connectivity index for a series of prodrug derivatives of 5-fluorouracil. The first order connectivity index correlated with the experimental logP values, molecular area and volume of all prodrugs. Comparison of ClogP values provided group contribution values while the comparison of molecular volume data provided information on the volumes of individual substituents relative to that of hydrogen.
Ren\textsuperscript{55}, in 2002, derived a novel vertex degree $v_m$ for heteroatom in molecular graph on the basis of the valence connectivity $v$ of Kier-Hall. The atom-type AI indices and Xu index, were modified for compounds with heteroatom by replacing the vertex-degree of hetero-atom by the proposed $v_m$. The modified Xu index and AI index provided QSPR models for the normal boiling points (BP). Molar volumes (MV)< molar refractions (MR) and molecular total surface areas (TSA) of alcohols with upto 17 non-hydrogen atoms. These physical properties were expressed as a linear combination of the individual indices related to molecular size and atom-type.

*Distance based graph invariants:*

These invariants employ distance matrix to characterize molecular graphs. Distance matrix $D(G)$, for a graph $G$ is defined as a real, square, symmetrical matrix of order $n$, with entries, $d_{ij}$, representing the distance traversed in moving from vertex $i$ to vertex $j$ in $G$. The $d_{ij}$ entries always satisfy the following criteria of identity relation, symmetry relation and triangle inequality:

\begin{align*}
    d_{ii} &= 0 \\
    d_{ij} &= d_{ji} \\
    d_{ij} &\leq d_{ik} + d_{kj}
\end{align*}

The entry $d_{ij}$ in the distance matrix $D(G)$ indicates the number of edges in the shortest path between vertices $i$ and $j$.

Wiener\textsuperscript{56}, in 1947, was first to propose graph invariant based on the distance matrix to calculate the boiling points. Wiener defined polarity number $p$ as the number of pairs of carbon atoms that are separated by three carbon-carbon bonds. The polarity number is therefore the sum of the entries of length three in the off-
diagonal matrix. He defined path number $W$ also known as Wiener's index as sum of distances between any two carbon atoms in the molecule in terms of carbon-carbon bonds. Wiener's index, which reflects the branching of the molecule, may be calculated from the sum of the off-diagonal elements of the distance matrix.

$$W(G) = \sqrt{\frac{1}{2} \sum_{i,j} d_{ij}}$$

*Where* $d_{ij}$ *is the number of edges in the shortest-path that connects the pair of atoms* $i$ *and* $j$.

The advantages of Wiener's index are:

1. Simple computation.
2. Can be employed for both cyclic and acyclic structures.
3. Is a biparametric index i.e. indicative of both size and shape of a molecule.
4. Correlates with physicochemical properties and biological activities.

Hosoya\textsuperscript{57}, in 1971, suggested for the first time a nontrivial single graph descriptor ($Z$), for expressing structure-property relationship. Although the Hosoya's $Z$ index has been associated with the adjacency matrix, it can be classified among the distance matrix invariants due to the procedure used to calculate $p(G,k)$

$$Z = \sum_{k} p(G,k)$$

*Where* $p(G,k)$ *represents the number of ways in which* $k$ *bonds are so chosen from graph* $G$ *that no two of them are adjacent.*

For linear graphs $G$, $Z$ can also be defined as the sum of the absolute values of coefficients in the characteristic polynomial.
\[ P_0(X) = \sum_{k=0}^{m} (-1)^k \times p(G, k) \times X^{N-k} \]

Where \( m \) is the largest number of bonds disconnected to each other in \( G \).

Z index is correlated well with the mode of branching and ring closure.

Hosoya\textsuperscript{58}, in 1972, proposed to use the Z index as a first sorting device for coding or retrieving the structure of the compounds with or without rings. He further recognized the problem that if the invariants of two given graphs are equal it is not sufficient for identity of structures, one must resort to finer comparison or identification. However, if Z values of two given graphs are different, one need not try further comparison. Hosoya’s Z index offers certain advantages over other invariants:

1. High discriminating power.
2. It can be computed from the molecular structure as well as polynomial.
3. It can be extended to matching polynomial and distance polynomial because it is defined through the counting polynomial, which is closely related to the characteristic polynomial.
4. It is capable of considering the influence of electronic systems.

Hosoya et al.\textsuperscript{59}, in 1975, proposed a generalization of Z index to account for the contributions made by unsaturated system in the structure of a molecule.

Hosoya’s Z index despite being discriminatory, possesses following limitations:

1. It represents a vague graph topological nature of molecular structure with respect to branching and cyclization.
2. No provision to consider heteroatoms.
3. Sophisticated programs are needed to compute Z index as the size, branching and compactness of a molecule increases.

Bonchev et al.\textsuperscript{60}, in 1980, developed total distance rank \((d_i)\), analogous with the vertex degree \(\delta_i\) proposed by Gutman et al., which is based on the distance sum \((V_{d_i})\) of the vertex \(i\) defined by the sum of the all entries \((d_{ij})\) of the \(i\)th row in the distance matrix \(D\).

Balaban\textsuperscript{61}, in 1983, described two mean distance based invariants \(- D \) and \(D1\) with modest discriminating ability and one average distance sum connectivity index \(J\) that was claimed to be least degenerate.

\[
D^i = \left( \frac{\sum g_i \times i^t}{\sum g_i} \right)^{1/k}
\]

\[
J = \frac{q}{\mu + 1} \sum s_{ij}^{-1} \frac{1}{\sqrt{s_i \times s_j}}
\]

Where \(g_i\) is the frequency of distance \(i\) in the graph \(G\), cyclomatic number \(\mu\) (number of rings in the graph) is \(\mu = q - n + 1\), \(q\) is the number of edges, \(s_i\) and \(s_j\) are distance sums.

Randic\textsuperscript{62}, in 1984, proposed identification number that was unique had structural significance and was easy to derive. Molecular ID number was defined as the count of all weighted paths in the structure. Molecular ID number has been found to be useful for the storage and comparison of chemical structures.

Kier\textsuperscript{63}, in 1985, developed molecular shape index, \(^2\kappa\), based on the count of the two-bond fragments. Kier introduced the term, \(^2\text{P}_{\text{min}}\), as the count of the number of two-bond paths, which for any non-cyclic molecule is equal to
\[ 2P_{\text{max}} = A - 2 \]

Where \( A \) is the total number of vertices in the graph.

On the other hand, \( 2P_{\text{max}} \) can be calculated for any isomeric series through

\[ 2P_{\text{max}} = \frac{(A-1)(A-2)}{2} \]

All non-cyclic molecules with in an isomeric series have \( 2P_i \) values at or within the bounds of the minimum and maximum value of \( P \)

\[ 2P_{\text{min}} \leq 2P_i \leq 2P_{\text{max}} \]

After appropriate considerations and refinements, Kier derived the final equation as per following:

\[ 2\kappa = \frac{2(2P_{\text{max}} - 2P_{\text{min}})}{2P_i^2} = \frac{(A-1)(A-2)^2}{2P_i^2} \]

According to Kier the \( 2\kappa \) index describes the molecular shape in relationship to the star and linear graph and is normalized to the number of vertices. The Kappa shape invariants are the basis of a method of molecular structure quantification in which attributes of molecular shape are encoded into three invariants (Kappa values). These Kappa values are derived from counts of one-bond, two-bond and three-bond fragments, each count being made relative to fragment counts in reference structures which possess a maximum and minimum value for that number of vertices.

Rouvray\(^{64}\), in 1986, reviewed the importance of graph invariants in chemistry and investigated the dependence of physicochemical properties such as viscosity, density and surface tension of alkanes, alkenes and alkynes on graph topological factor employing Wiener's index.
However Wiener's index is susceptible to following limitations:

1. High degree of degeneracy, as a result characteristic index values are not obtained for each molecule.

2. Heteroatoms and unsaturated bonds in molecular structures are treated in a manner similar to carbon atoms and single bonds respectively during the computation of index.

3. Though combinatorial origin of the index is evident, its interpretation is less apparent.

Despite these limitations Wiener's index has been found to be useful for the prediction of activity of diverse bioactive molecules. The role of Wiener's index in SAR of bioactive molecules have been explored in areas such as toxicity of alcohols, analgesic property of ketobemidones, herbicidal activity of triazinone and narcotic property of barbiturates\textsuperscript{61}.

Balaban\textsuperscript{65}, in 1986, reported a special treatment for unsaturated system of molecules employing fractional distances in the distance matrix. Balaban adapted average distance-sum-connectivity index, J, similar to Randic's index for heteroatoms based on their electronegativities. Balaban's J index possesses high discriminating power.

Schultz\textsuperscript{66}, in 1989, introduced a graph invariant termed as molecular topological index – MTI(V) for the characterization of alkanes with respect to their boiling points.

Muller \textit{et al.}\textsuperscript{67}, in 1990, systematically tested MTI for alkane trees up to 16 carbon atoms and reported counter examples. They also tried to redefine MTI in terms of distance matrix and adjacency matrix but were not successful. MTI is liable to following limitations:
1. Low discriminating power is evident when two isomorphic alkane trees of different sizes yield same value of the index.

2. The computation of the index involves the use of adjacency matrix that lacks distinction between molecular topological distances of length 1 and 2.

3. It is not suited for cyclic structures.

**Hall and Kier**\(^{68}\), in 1990, defined topological state invariants as numerical values associated with each vertex in a molecule and capable of encoding information about the molecular topological environment of that vertex due to all other vertices in the molecule. The molecular topological relationship to each other vertex is based on the encoding of vertex information in all the paths emanating from that vertex. Molecular topologically equivalent vertices have identical values of the molecular topological state index \(S\), and inequivalent vertices have different values.

**Petitjean**\(^{69}\), in 1992, defined the shape coefficient \(I\) of chemical compound as the ratio \((D-R)/R\) where \(R\) is the generalized radius and \(D\) is the generalized diameter.

The graph radius \((R)\) is defined as the smallest vertex eccentricity in the graph. The graph diameter \((D)\) is defined as the largest vertex eccentricity in the graph. The radius-diameter diagrams may then be used as vertex graph invariants.

**Ivanciuc et al.**\(^{70}\), in 1993, reported that the graph distance matrix for a tree contains the same information as in the adjacency matrix. Its use in applications in chemistry perhaps is limited because the entries in such a matrix for more distance vertices increase while interactions of distant atom decrease. Therefore distance matrix can be modified, by considering reciprocal distance \(1/D\) (as in potential, reciprocal \(1/D^2\) (as in force), and other similar constructions, like \(1/D^6-1/D^{12}\) (as in the Lennard-Jones long-range potential). Thus modified matrices may better simulate
interactions between atoms at greater distances and may produce more suitable mathematical descriptors.

Randic et al.\textsuperscript{71}, in 1993, constructed new matrices by generalizing Wiener's procedure and reported the sequences (higher Wiener numbers, $W$) generated by summing the entries in the matrix for vertices at the same distance from one another. Wiener matrix was extended to consider heteroatoms.

Randic\textsuperscript{72}, in 1993, defined hyper Wiener index (WW), modifications of Wiener index, for any connected acyclic structure as a sum of entries in the sequence.

Randic et al.\textsuperscript{73}, in 1994, reported the first eigen value of Wiener matrix as an alternative index for the molecular branching. JJ index was developed in analogy with the connectivity index from the adjacency matrix. First row sums, $R_i$, were obtained and these numbers were used in formula $1/(R_iR_j)^{1/2}$ for all pairs of adjacent vertices.

Randic\textsuperscript{74}, in 1994, introduced a novel matrix associated with molecular graphs, the construction of which was related to Hosoya index. He also proposed two new invariants $1Z$ and $2Z$ for structure-property and structure-activity studies.

\[ Z = \sum_{i<j} Z_{ij} \]

Where summation goes over all off-diagonal entries in the upper triangle of the \(^*Z\) matrix.

Chang and Lu\textsuperscript{75}, in 1994, provided counter examples to prove that topological state method cannot determine classes of symmetrically equivalent atom in a molecule and total topological index is highly discriminating but not unique.

Galvez et al.\textsuperscript{76}, in 1994, reported charge invariants $G_k$ and $J_k$. Their ability for the description of molecular charge distribution was established with the help of
dipole moment. Since charge influence decreases with square of the distance 'inverse square distance matrix' was used as first charge descriptor.

\[ G_k = \sum_{i=1}^{N} \sum_{j=i+1}^{N} |CT_i| \delta(k, D_{ij}) \]

\[ J_k = \frac{G_k}{N-1} \]

Where \( N \) is number of vertices in the graph, \( CT \) is charge term, \( D_{ij} \) are entries of the distance matrix and \( k \) is index order of \( G \). Therefore, \( G_k \) evaluates total charge transfer between atoms placed at a distance \( k \) and \( J_k \) represents mean value for the charge transfer for each bond.

Khadikar et al., in 1995, introduced Szeged index \( S_z \) analogous to Wiener's index but not of the same type. Whereas Wiener index involves properties of pairs of vertices that are at large distances, the \( S_z \) index deals with properties of pairs of vertices at unit distance. Wiener's index is globally defined whereas \( S_z \) index locally defined.

Estrada and Gutman, in 1996, proposed molecular topological index \( MTI(E) \) based on edge-distances in molecular graphs. \( MTI(E) \) was defined in analogy to the graph invariant of Schultz.

De Gregorio et al., in 1998, made a structure-activity analysis of a series of steroids binding to corticosteroid-binding globulin using the electrotopological state index for each atom in the molecule. Two indices were found to correlate with the binding affinity. These indices encoded structural characteristics in the ‘A’ and the ‘D’ rings of the steroids in the study. One of the indices was formulated as the
difference between two indices in the 'A' ring. The influence of structure changes in or near the 'A' ring could be monitored by composite index.

Chan et al.\textsuperscript{80}, in 1998, showed that there exist algebraic connections between various molecular graph based descriptors. Hosoya index, Narumi's simple topological index (product of all vertex degrees) and connectivity index of Randic which were believed not to be related to \(W\), exhibit algebraic connections with \(W\). Such connections can be avoided if instead of molecular graph appropriately weighted variants are used.

Gupta et al.\textsuperscript{81}, in 1999, conceptualized a novel topological descriptor termed as superpendentic index and investigated its discriminating power with regard to antiulcer activity of 4-substituted-2-guanidino thiazoles derivatives.

Espeso et al.\textsuperscript{82}, in 2000, proposed decomposition of Hosoya matrix into the sum of \(\chi Z\) matrices. The Hosoya hyperindex \(H\) and a set of invariants, \(\chi Z\), were proposed.

Galvez et al.\textsuperscript{83}, in 2000, introduced invariants of differences of path lengths (DPs), and demonstrated that the overall sum of the inverse of the squares of the differences of distances between all pairs of vertices of the graph is a measure of the mean global kinetic energy of the electrons. These invariants were employed to predict the resonance energies and in the evaluation of biological properties such as antibacterial activities of a wide set of heterogeneous compounds.

Balaban et al.\textsuperscript{84}, in 2000, obtained reverse Wiener matrix by subtracting from the graph diameter all topological distances. The half sum of sums over rows was reported as reverse Wiener (RW) index.

Randic and Zupan\textsuperscript{85}, in 2001, suggested that invariants should be discussed in terms of their partitioning into bond contributions, which for different invariants
and different bonds will assume different values. The Wiener index was modified to $W^*$ in which bond contributions were determined using the reciprocal of the product of the number of atoms on each side of a bond. Similarly, Hosoya index $Z$ was modified to $Z^*$ by considering the frequency of occurrence of CC bonds in the patterns of disjoint bonds. Finally, invariants were classified into two groups depending on the relative magnitudes of the terminal and internal CC bond contributions in the partitioning of the invariants into bond contributions.

Ivanciuc et al.\cite{Ivanciuc}, in 2001, introduced a new set of invariants ($W_{e\circ o}, W_{o\circ e}, W_{0\circ 0}, W_{e\circ 0}$) representing a partitioning of the Wiener index based on the counts of even and odd molecular graph distances. The invariants were generalized by weighing exponents that can be optimized during structure activity/property relationship modeling process.

Balaban et al.\cite{Balaban}, in 2001, derived a general formula $W^\sim = d/[3(v+0)]$ for the normalized Wiener index of polymers. It made possible the calculation of the graph invariant directly from simple structural information: the number of atoms ($v$) and rings ($0$) in the repeating polymer cell, and the topological distance $d$ between the corresponding pairs of equivalent atoms in two neighboring monomer units. The reciprocal relationship with the similar index $J^\sim$ was pointed out and an approximate hyperbolic dependence is presented between these two indices.

Estrada and Molina\cite{Estrada}, in 2001, defined novel molecular descriptors based on local spectral moments of the bond matrix. Mathematical expressions relating bond moments to linear combinations of structural fragments were derived. The novel descriptors were used to predict boiling points of alcohols and specific rate constant (log $k$) of the nucleophilic addition of mercaptoaetic acid to 2- furylethylene derivatives.
Torrens, in 2001, elucidated polar character of hydrocarbons using topological charge-transfer index. The new charge-transfer index elucidates whether hydrocarbons are non-polar.

Khadikar et al., in 2001, developed Padmakar-Ivan (PI) index of the molecular graphs of unbranched catacondensed benzenoid hydrocarbons. The PI index for hexagonal chains is defined as summation of sums of edges of n\text{_{eu}} and n\text{_{ev}} over all the edges of connected graph G. Where n\text{_{eu}} is the number of edges of G lying closer to u than to v and n\text{_{ev}} is the number of edges lying closer to v than to u.

Khadikar et al., in 2002, developed quantitative structure-toxicity relationship (QSTR) to predict hydrophobicity (log P), toxicity of benzene derivatives using Padmakar-Ivan (PI) index. The predictive ability of the multiparametric models was cross-validated.

Hou, in 2002, characterized trees with a given size of matching and having minimal and second minimal Hosoya index. The Hosoya index of a graph is defined as the total number of independent edge subsets of the graph.

Quigley and Naughton, in 2002, investigated intercorrelation between a series of physicochemical parameters and topological indices for a set of \beta-blockers. A novel index eccentric adjacency index proposed by Gupta et al, newly developed descriptor valence eccentric adjacency index along with distance based indices were included in the study. A high degree of linear correlation between the connectivity indices was noted.
Centric graph invariants:

The concept of graph center is based on molecular topological distances between the graph vertices. The center vertices have the smallest maximal distance to the other vertices. The distance between vertices is defined as:

\[ d_{ij} = \min \text{ for } j = 1, 2, \ldots, p \]

Invariants derived from the concept of center are called centric graph descriptors.

Balaban, in 1979, proposed a set of five graph invariants classified as centric invariants on the basis of sequences of numbers obtained by pruning an acyclic graph. By pruning stepwise all vertices of degree one (δ), a vertex (center) or an edge connecting two adjacent vertices (bicenter) is obtained. Balaban developed a centric index B for branching.

\[ B = \sum \delta_i^2 \]

Four invariants were devised from B and M1 to differentiate branching from number of vertices by normalization or binormalization. Normalized and binormalized centric (C, C') and quadratic invariants (Q, Q') were defined as

\[ C = \frac{B - 2n + U}{2} \]

\[ C' = \frac{B - 2n + U}{(n-2)^2 - 2 + U} \]

\[ Q = 3V_4 + V_1 \]

\[ Q' = \frac{2(3V_4 + V_3)}{(n-2)(n-3)} \]
Where \( n \) is the number of vertices, 
\[
U = \frac{1 - (-1)^n}{2}
\]
while \( V_3 \) and \( V_4 \) are vertices of degree three and four respectively.

It was found that the centric invariants parallel the ordering induced by index \( B \), while the quadratic invariants induced ordering which parallel those due to Gutman et al's index \( M_1 \) and Gordon-Scantlebury index \( N_2 \).

**Bonchev et al.**\(^9^6\), in 1980, generalized the concept of center for any connected acyclic or cyclic graph on the basis of topological distance matrix using four criteria in a specified order. The center points provide the smallest distance to any other vertex; the smallest sum of distances to all other vertices; the smallest number of times the maximum distance occurs in the distance code and constancy on repeating above three criteria to the pseudo center graph containing only the vertices selected by the previous criteria and their incident edges. The centric invariants were proposed for correlations, for estimating the shape of isomeric chemical structures and for use in coding and computer processing of chemical structures.

**Diudea et al.**\(^9^7\), in 1991, introduced the so-called \( B \)-matrix for the development of a new index by counting the vertexes in graph spheres (layers). A sphere is a list of atoms surrounding a central vertex in a given topological distance. Using spheres is advantageous in studies that investigate the influence of neighbors on a specific property of central vertex.

**Diudea**\(^9^8\), in 1994, distinguished the sequence matrices (SM) and layer matrices (LM). A sequence matrix is a collection of walks starting from the vertices \( i \) to all other \( n-1 \) vertices in \( G \). While a layer matrix collects the properties of vertices \( u \) located in concentric shells (layers) at distance \( j \) around each vertex \( i \) in \( G \). Further he defined two invariants based on LM matrices – centrocomplexity \( x(LM) \) and centric invariant \( c(LM) \)
Where $m$ is the label for a particular type of walk or property, $n$ the number of vertices in $G$, $d$ is the diameter of $G$ and $esp$ being eccentricity of vertex $i$.

**Information theory based graph invariants:**

Information theory can be applied to determine structural information – the quantity of information contained in the structure of a system. For a system having $N$ elements partitioned into subsets of equivalent elements $N_1, N_2, \ldots, N_k$, according to Shannon information theory, if a probability can be associated $p_1, p_2, \ldots, p_k$, then probability distribution of the elements is the quantity of information. The mean

$$I = \sum_{i=1}^{k} p_i \times \log_2 p_i$$

information contained in each element of the structure can be defined as

*Where $p_i$ is the probability of randomly selected element. The information content in the entire molecule with $N$ vertices is given by $N$.*

**Bonchev and Trinajstic** in 1977, used information theory for defining information for adjacency, incidence, polynomial coefficients of the adjacency matrix and for distances of molecular graph.
\[
I_{\text{ad}} = N^2 \log_2 N^2 - 2(N - 1) \log_2 2(N - 1) - \left(N^2 - 2N + 2\right) \log_2 (N^2 - 2N + 2)
\]

\[
I_{\text{nc}} = N(N - 1) \log_2 N(N - 1) - 2(N - 1) \log_2 2(N - 1) - (N - 1)(N - 2) \log_2 (N - 1)(N - 2)
\]

\[
I_{\text{pc}} = Z \log_2 Z - \sum_{k=0}^{N^2/2-1} p(G,k) \log_2 p(G,k)
\]

\[
I_{\text{ac}} = N^2 \log_2 N^3 - N \log_2 N - \sum_{i=1}^{n} 2k, \log_2 2k,
\]

Where \( N \) is the number of vertices, \( Z \) is Hosoya's index, \( p(G,k) \) is probability of randomly chosen polynomial coefficient, \( 2k \) is the number of times the distance value \( i \) appears in the distance matrix.

These invariants are largely defined from a combination of the parameters used to obtain Wiener \( W \) index, the Hosoya \( Z \) index and the Randic connectivity index.

Bertz\(^{100}\), in 1981, proposed a new general information index that incorporates the information on atomic composition, information on graph connections and molecular size. The general index of molecular complexity of Bertz is given by the equation

\[
I_{\text{Bertz}} = I_{\text{ac}} + I_{\text{CONN}} + I_{\text{SIZE}}
\]

Chirality and stereochemistry can be reflected by \( I_{\text{CONN}} \) the distribution of connections into classes of orbital equivalency.

Raychaudhury et al.\(^{101}\), in 1984, defined three information invariants - degree complexity \( (I^D) \), graph vertex complexity \( (H^V) \) and graph distance complexity \( (H^D) \). Graph distance complexity was found to be the only index to discriminate all the graphs studied.
Klopman and Raychaudhury, in 1988, described an information index (vertex distance complexity, $V^d$) for the vertices of a molecular graph and used the same for qualitative evaluation of mutagenic activity of a series of nonfused ring aromatic compounds.

According to Randic, in 1991, a necessary step in the development of any graph invariant for a chemical structure is to include the differentiation of heteroatoms present in the molecules. The neglect of heteroatoms and spatial features of molecules has been considered by as severe limitations of graph models. Moreover, to consider a graph invariant as a molecular descriptor it needs to have several desirable attributes. For practical purposes it is convenient that candidates to molecular descriptors have good correlations with at least one physical property.

Randic et al., in 1994, introduced distance/distance matrices for graphs, embedded on two- and three-dimensional grids. The first eigen value of these matrices, $\lambda/n$ for path graphs was reported to be index of folding. The ratio $\phi = \lambda/n$ approaches 1 for geometrically linear structures while it approaches 0 as path graph is repeatedly folded.

Amic and Trinajstic, in 1995, defined the Detour matrix $\Delta = \Delta(G)$ of a labelled connected graph $G$ as a real symmetric $N$ by $N$ matrix whose $(i,j)$ entry is the length of the longest path from vertex $i$ to vertex $j$. This definition is opposite to the traditional distance matrix.

$$\Delta = \sum_{i=0}^{l_{max}} B_i$$

Where $l_{max}$ is the distance of the longest possible path in a graph $G$, while $B_i$ is an auxiliary matrix defined as
\[ B_i = \begin{cases} 0 & \text{if path exists between } i \text{ and } j \text{ of length less than or equal to } l \\ 1 & \text{otherwise} \end{cases} \]

Randic et al.\textsuperscript{106}, in 1997, reported a new route to construct structural invariants for molecules by associating a matrix M(S), the elements of which were qualified molecular substructures. Selecting a matrix invariant of M and a structure invariant for the substructures that represent elements of M derived the new invariants, called double invariants.

Randic et al.\textsuperscript{107}, in 2000, reported partial ordering of folded structures by considering regularities for the leading eigen values of the D/D matrices and the leading eigen values of the line-adjacency matrices selected folded structures. The line adjacency matrix is of interest in numerical characterization of DNA primary sequences.

Nachbar\textsuperscript{108}, in 2000, developed a simple hierarchical data structure (tree) and associated set of algorithms that permitted the direct manipulation of the topology of a molecule while maintaining valid chemical valence. The performance of the method was strongly dependent on the size and complexity of the evolved chemical structures.

Estrada\textsuperscript{109}, in 2001, developed a generalized molecular graph matrix. An invariant defined on the basis of this matrix generalizes some of the classical topological indices. Thus, Wiener index, Zagreb group parameters, Balaban J index, Harary H number, Randic index and valence connectivity index were part of an infinite set of molecular descriptors that could be derived from the same invariant.

Miscellaneous Graph Invariants

Owing to the loss of information resulting from the condensation of graph features into a single number, none of the known graph invariants can uniquely
characterize molecular graphs. Despite this limitation graphical approach has found applications in diverse areas of scientific research such as engineering, computer science, geography, architecture, physics, chemistry, toxicology, pharmacology, and pharmaceutical chemistry.

Bonchev et al.\textsuperscript{110}, in 1992, utilized graph topological extrapolation method for the modeling of polymer properties (TEMPO) that was based on the graph topological description of the polymer elementary units by means of the normalized Wiener number represented as a polynomial of degree 3 with respect to the number of atoms. The method was applied to the calculation of p-electron energies and energy gaps of various conjugated polymers, as well as to the assessment of the melting point, density, refractive index, and specific rotation of some industrially produced polymers.

Graph invariants developed for predicting physicochemical properties and biological activities, of chemical substances, have also been used for drug design.

Mekenyan et al.\textsuperscript{111}, in 1993, applied optimized approach based on Structural Index Sets to model the bronchospasmolytic activity and toxicity of theophylline derivatives. The opposite influence of graph topology on bronchospasmolytic activity and toxicity was established.

Gealy et al.\textsuperscript{112}, in 1996, compiled a database of chemicals causing allergic contact dermatitis (ACD) to discern structure-activity relationships. Clinical reports represented a considerable fraction of the data. The model was used to predict the activity of chemicals identified from clinical case reports.

Sharma et al.\textsuperscript{113}, in 1997, reported a novel adjacency cum distance cum topological index termed as eccentric connectivity index. This index was successfully employed by the authors for the development of a model for prediction of analgesic
activity of substituted piperidinyl Me ester and methylene Me ester analogues. This index was also found to be highly beneficial for structure property relationship.

DeJulian-Ortiz et al.\textsuperscript{114}, in 1998, proposed chiral topological invariants that can differentiate pharmacological activity between pairs of enantiomers. Chiral behavior is not associated with 3-dimensional geometry but is related to symmetry which allowed handling of chiral atoms by weighted graphs. The 50% inhibitory concentration values of D\textsubscript{2} dopamine receptor and the \(\sigma\) receptor for a group of 3-hydroxy phenyl piperidines were predicted.

Borodina et al.\textsuperscript{115}, in 1998 applied a method based on topolelectrical invariants to estimate the synthetic molecule resemblance to small endogenous bioregulators. The results demonstrated discriminative ability of proposed structure description and measure of similarity.

Cercos-Del-Pozo et al.\textsuperscript{116}, in 1999, designed new compounds showing hypolipemic activity using computer-aided method based on molecular topology and QSAR analysis. \textit{Connectivity functions were used to design three drugs that were tested for hypolipemic property in rats.}

Golbraikh et al.\textsuperscript{117}, in 2001, introduced several series of novel chirality descriptors of chemical organic molecules. These included modified molecular connectivity indices, Zagreb group indices, extended connectivity and overall connectivity, and topological charge indices. These modified descriptors made use of an additional term called chirality correction, which was added to the vertex degrees of asymmetric atoms in a molecular graph. These novel chirality descriptors were implemented in a QSAR study of a set of ecdysteroids with a high content of chiral and enantiomeric compounds. The results compared favorably with those obtained with the comparative molecular field analysis.
Kauffman\textsuperscript{118}, in 2001, used experimental IC\textsubscript{50} data for 314 selective cyclooxygenase-2 (COX-2) inhibitors to develop quantitation and classification models as a potential screening mechanism for larger libraries of target compounds. Numerical descriptors encoding solely topological information were calculated for all structures and were used as inputs for linear regression, computational neural network and classification of analysis routines. High degree of predictions were obtained.

Mattioni\textsuperscript{119}, in 2002, developed mathematical models to find QSAR relationships that correlate chemical structure and inhibition towards carbonic anhydrase isoenzymes. Good prediction rates were obtained with these models.

Golbraikh \textit{et al.}\textsuperscript{120}, in 2002, introduced several series of novel ZE-isomerism descriptors derived directly from two dimension molecular topology. The applicability of ZE - isomerism descriptors to QSAR analysis was demonstrated in the studies of a series of 131 anticancer agents inhibiting tubulin polymerization.

Gupta \textit{et al.}\textsuperscript{121}, in 2002, reported eccentric distance sum, a novel graph invariant to investigate its potential in structure activity / property relationship. The SAR of eccentric distance sum was investigated with regard to anti HIV activity of dihydroseselins. The accuracy of prediction was found to be more than 88\% with regard to anti HIV activity. The investigations pertaining to QSPR with regard to various physical properties of primary amines, secondary amines and alcohols revealed correlation percentages ranging from 93 – 99 \%. The overall results with regard to SAR and QSPR using eccentric distance sum were better than the corresponding values obtained using Wiener’s index.

Review of literature reveals that though a very large number of graph invariants have been reported but only a handful of them have been successfully
employed in structure activity relationships. Accordingly there is a strong need for developing novel graph invariants with high discriminating power but without any degeneracy so as to facilitate ease in development of models for prediction of biological activities of diverse nature.