



■ OASIS method (\equiv *Optimized Approach based on Structural Indices Set*)

The OASIS method is based on the same assumptions as \rightarrow *Hansch analysis* and can be regarded to as an extended and optimized version of the Hansch approach [Mekenyan and Bonchev, 1986; Mekenyan, Karabunarliev *et al.*, 1990a].

Besides \rightarrow *substituent constants*, the descriptors used in the OASIS approach are \rightarrow *topological indices*, \rightarrow *geometrical descriptors*, \rightarrow *steric descriptors*, and \rightarrow *electronic descriptors*, both for molecules and their fragments.

The OASIS model is defined as:

$$\text{biological activity} = f(\{\text{TI}\}, \{\Phi_1\}, \{\Phi_2\}, \{\Phi_3\})$$

where $\{\Phi_i\}$ are different sets of \rightarrow *physico-chemical properties*, each set representing steric, electronic, and hydrophobic descriptors, respectively, and $\{\text{TI}\}$ is a set of topological descriptors.

This equation allows for the possibility that more than one descriptor of the different factors contribute to the overall biological activity. Moreover, these model components can be either local, that is, referring to atoms or fragments, or global, that is, describing the molecule as a whole.

📖 [Mekenyan, Bonchev *et al.*, 1986, 1988b; Mekenyan, Peitchev *et al.*, 1986a, 1986b; Mekenyan, Karabunarliev *et al.*, 1990b; Mercier, Mekenyan *et al.*, 1991; Bonchev, Mountain *et al.*, 1993; Mekenyan, Mercier *et al.*, 1993; Bonchev, Seitz *et al.*, 1994; Kamenska, Mekenyan *et al.*, 1996]

- **object** \rightarrow data set
- **occupancy numbers** \rightarrow cell-based methods
- **Ochiai similarity coefficient** \equiv *cosine similarity coefficient* \rightarrow similarity/diversity (☉ Table S9)
- **octanol–water distribution coefficient** \rightarrow physico-chemical descriptors (☉ partition coefficients)
- **octanol–water partition coefficient** \rightarrow physico-chemical descriptors (☉ partition coefficients)
- **octupole moment** \rightarrow electric polarization descriptors
- **OCWLI** \equiv *Optimization of Correlation Weights of Local Invariants* \rightarrow variable descriptors
- **odd–even index** \rightarrow Cao–Yuan indices
- **OEI** \equiv *odd–even index* \rightarrow Cao–Yuan indices
- **oligocenter** \rightarrow center of a graph

■ Omega polynomial

The Omega polynomial is a \rightarrow counting polynomial based on the counting of the so-called “quasiorthogonal cuts” (*qoc*) of a graph as

$$\Omega(G; x) = \sum_c m(G; c) \cdot x^c$$

where the coefficients $m(G; c)$ are the numbers of occurrences of quasiorthogonal cuts of length c (i.e., the number of edges cutoff) and the summation runs over the maximal length of *qoc* in the graph [Diudea, 2006; Diudea, Vizitiu *et al.*, 2007; Diudea, Cigher *et al.*, 2008].

The quasiorthogonal cuts are defined as the following. Two edges $e(u, v)$ and $e'(u', v')$ are called *codistant* (e co e') if for $k = 0, 1, 2, \dots$ there exist the relationships

$$d(u, u') = d(v, v') = k \quad \text{and} \quad d(u, v') = d(v, u') = k + 1$$

or *vice versa*. For some edges of a connected graph the following relationships may exist:

1. e co e'
2. e co $e' \Leftrightarrow e' \text{ co } e$
3. e co $e' \quad \wedge \quad e$ co $e'' \Leftrightarrow e' \text{ co } e''$

although the third relationship is not always valid.

Now, let $C(e)$ denote the set of all edges of the graph that are codistant to the edge e , that is, equidistant and “topologically” parallel. If all the elements of $C(e)$ satisfy the relationships 2 and 3, the $C(e)$ is called *orthogonal cut* (*oc*) of the graph and the graph is called *co-graph* if and only if the set \mathcal{E} of all edges in the graph is the union of disjoint orthogonal cuts:

$$C_1 \cup C_2 \cup \dots \cup C_k = \mathcal{E} \quad \wedge \quad C_i \cap C_j = \emptyset \quad \text{for } i \neq j \text{ and } i, j = 1, 2, \dots, k$$

If any two consecutive edges of a cut edge sequence are codistant (i.e., obeying the relations 2 and 3) and belong to one and the same face of the covering, such a sequence is called a *quasiorthogonal cut* (*qoc*) strip. A *qoc* strip starts and ends either out of G (at an edge with end points of degree lower than 3, if G is an open lattice) or in the same starting polygon (if G is a closed lattice).

From the Omega polynomial, two molecular descriptors are derived. The first one, called **Cluj-Ilmenau index**, denoted as CI , is derived from the first Ω^I and second Ω^{II} derivatives of the Omega polynomial as

$$CI = (\Omega^I)^2 - (\Omega^I + \Omega^{II}) \quad \text{at } x = 1$$

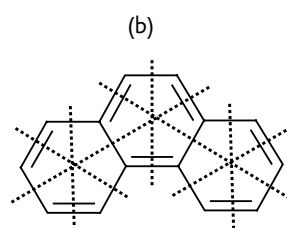
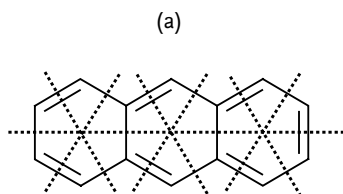
This index coincides with the \rightarrow *PI index* for polycyclic graphs.

The second index, denoted as I_Ω , is calculated from the summation of all the possible Omega polynomial derivatives Ω^d at $x = 1$, and normalized to the first polynomial derivative (which is equal to the number of edges in the graph):

$$I_\Omega = \frac{1}{\Omega^I} \cdot \sum_d (\Omega^d)^{1/d} \quad \text{at } x = 1$$

Example O1

Omega polynomials, their derivatives, and CI and I_Ω indices for anthracene (a) and phenanthrene (b). PI is the PI index. Straight line segments indicate orthogonal edge cuts.



$$\Omega(a) = 6 \cdot x^2 + x^4$$

$$\Omega^I(a) = 12 \cdot x + 4 \cdot x^3 \quad \Omega^{II}(a) = 12 + 12 \cdot x^2$$

$$\Omega^{III}(a) = 24 \cdot x \quad \Omega^{IV}(a) = 24$$

$$CI(a) = 256 - (16 + 24) = 216 = PI(a)$$

$$I_\Omega(a) = \frac{1}{16} \cdot (16 + 24^{1/2} + 24^{1/3} + 24^{1/4}) = 1.624$$

$$\Omega(b) = 5 \cdot x^2 + 2 \cdot x^3$$

$$\Omega^I(b) = 10 \cdot x + 6 \cdot x^2 \quad \Omega^{II}(b) = 10 + 12 \cdot x$$

$$\Omega^{III}(b) = 12 \quad \Omega^{IV}(b) = 0$$

$$CI(b) = 256 - (16 + 22) = 218 = PI(b)$$

$$I_\Omega(b) = \frac{1}{16} \cdot (16 + 22^{1/2} + 12^{1/3}) = 1.436$$

The **non-Omega polynomial** is derived from the Omega polynomial so that to be a complementary quantity as

$$N\Omega(G; x) = \sum_c m(G; c) c x^{(B-c)}$$

where the coefficients $m(G; c)$ are the occurrences of quasiorthogonal cuts of length c and the summation runs over the maximum length of qoc in the graph; B is the total number of edges in the graph [Diudea, Vizitru *et al.*, 2007]. The first derivative of the non-Omega polynomial at $x = 1$ coincides with the $\rightarrow PI$ index.

- **optimal descriptors** \equiv *variable descriptors*
- **optimization** \rightarrow chemometrics
- **Optimization of Correlation Weights of Local Invariants** \rightarrow variable descriptors
- **Optimized Approach based on Structural Indices Set** \equiv *OASIS method*
- **orbital electronegativity** \rightarrow quantum-chemical descriptors (\odot electronic chemical potential)

■ **orbital information indices** (\bar{I}_{ORB})

The first information index proposed as a measure of complexity of a \rightarrow *H-depleted molecular graph* is the **vertex orbital information content** \bar{I}_{ORB} , originally simply called **topological information content** \bar{I}_{TOP} [Rashevsky, 1955; Trucco, 1956a, 1956b], defined as \rightarrow *mean information content*:

$$\bar{I}_{ORB} \equiv \bar{I}_{TOP} = - \sum_{g=1}^G \frac{A_g}{A} \cdot \log_2 \frac{A_g}{A}$$

where A is the total number of vertices and A_g is the number of topologically equivalent vertices of g th type; the summation runs over all G different classes of topological equivalence. The atoms are distinguished not on the basis of their chemical nature but on their topological relationships to each other. Vertices are topologically equivalent if they belong to the same \rightarrow *orbits* of the vertex \rightarrow *automorphism group* of the graph. Note that in the original definition of the topological information content proposed by Rashevsky, the equivalence classes were defined by the \rightarrow *vertex degree* of the neighboring vertices instead of the graph orbits.

The **total topological information content** was defined as

$$I_{TOP} = A \cdot \bar{I}_{TOP} = A \cdot \log_2 A - \sum_{g=1}^G A_g \cdot \log_2 A_g$$

where A is the number of vertices in the graph.

The **edge orbital information content** was defined by analogy as [Trucco, 1956a, 1956b]

$${}^E\bar{I}_{ORB} = - \sum_{g=1}^G \frac{B_g}{B} \cdot \log_2 \frac{B_g}{B}$$

where B is the total number of edges in the graph and B_g is the number of edges belonging to the g th edge orbit of the graph; the summation runs over all G different edge orbits. The corresponding **total edge orbital information content** was defined as:

$${}^E I_{ORB} = B \cdot {}^E\bar{I}_{ORB} = B \cdot \log_2 B - \sum_{g=1}^G B_g \cdot \log_2 B_g$$

Moreover, based on the connection orbits of the graph the **connection orbital information content** was also defined as [Bonchev, 1983]

$${}^{\text{CONN}}\bar{I}_{ORB} = - \sum_{g=1}^G \frac{(N_2)_g}{N_2} \cdot \log_2 \frac{(N_2)_g}{N_2}$$

where N_2 is the \rightarrow *connection number* of the graph and $(N_2)_g$ is the number of connections belonging to the g th orbit of the graph; the summation runs over all G different connection orbits. The corresponding **total connection orbital information content**, defined as

$${}^{\text{CONN}}I_{ORB} = N_2 \cdot {}^{\text{CONN}}\bar{I}_{ORB} = N_2 \cdot \log_2 N_2 - \sum_{g=1}^G (N_2)_g \cdot \log_2 (N_2)_g$$

is a component of the \rightarrow *Bertz complexity index* when calculated for \rightarrow *multigraphs*.

The orbital information indices can also be calculated for multigraphs, giving a higher measure of graph complexity as the multiplicity of the edges provides more graph orbits than the simple graph. In the case of graph vertices of the same chemical element, the orbital information content for multigraph coincides with the \rightarrow *neighborhood information content* of maximal order calculated for the H-depleted molecular graph:

$$\bar{I}_{ORB} = IC_{\max}$$

- **orbital interaction graph of linked atoms** → determinant-based descriptors (\odot general a_N -index)
- **orbital interaction matrix of linked atoms** → determinant-based descriptors (\odot general a_N -index)
- **orbits** → graph
- **ordered structural code** → self-returning walk counts
- **ordered walk count molecular code** → walk counts
- **order of a subgraph** → molecular graph
- **order of neighborhood** → indices of neighborhood symmetry
- **order parameter** → polymer descriptors
- **oriented graph** → graph (\odot digraph)
- **orthogonal descriptors** \equiv *orthogonalized descriptors*

■ **orthogonalized descriptors** (\equiv *orthogonal descriptors*)

These are obtained by applying an orthogonalization procedure to a selected set of → *molecular descriptors*. A descriptor X_j is made orthogonal to another descriptor X_i simply by regressing it against X_i and using as the new orthogonal descriptor the residual $X_j - \hat{X}_j$, where \hat{X}_j is the value of X_j calculated by the regression model. The residual, that is, the orthogonalized descriptor, represents the part of descriptor X_j not explained by the descriptor X_i ; the symbol ${}^i\Omega_j$ is usually used to indicate that descriptor X_j is made orthogonal to descriptor X_i . The orthogonalization process is an iterative procedure until the last considered descriptor is orthogonalized against the preceding orthogonalized descriptors [Randić, 1991b, 1991e, 1991f].

The orthogonalization procedure requires a prior ordering of the descriptors to which other descriptors are subsequently made orthogonal, that is, it requires → *basis descriptors*. Thus, it is evident that different orthogonalized descriptor bases derive from different ordering. However, in the case of path counts, connectivity indices, → *uniform length descriptors*, that are naturally ordered descriptors the orthogonalization procedure can be applied easily and the symbol Ω_j used simply for orthogonalized descriptors.

The most known orthogonalization technique is the Gram–Schmidt orthogonalization scheme [Golub and van Loan, 1983]. When a subset of descriptors has the same partial ordering they are simultaneously and mutually orthogonalized within the set itself after sequential orthogonalization to the preceding descriptors [Klein, Randić *et al.*, 1997]. Canonical orthogonalization, symmetrical orthogonalization, and optimal orthogonalization are techniques to perform this task.

To better explain this procedure a sequence of → *path counts* mP is considered. **Orthogonalized path counts** are calculated through the following steps: (a) The first path count 1P in the sequence is chosen as the first orthogonalized descriptor Ω_1 ; it can also be decomposed into the mean contribution represented by Ω_0 and the deviation from the mean represented by Ω_1 . (b) The second orthogonalized descriptor Ω_2 is calculated as a residual of the regression between the second path count 2P against the first one 1P ; the residual contains information independent of 2P . (c) The orthogonalization continues by regressing the path number 3P against the first path count 1P , and then the residual of this regression against the orthogonalized descriptor Ω_2 ; the residual of the last regression is the third orthogonalized descriptor Ω_3 . In general, the p th orthogonalized descriptor Ω_p is defined as the residual in the multiple regression of the p th descriptor of the sequence against the $p - 1$ previously orthogonalized descriptors [Šoškić, Plavšić *et al.*, 1996b].

Orthogonalized descriptors are used in \rightarrow *similarity/diversity* analysis and quantitative \rightarrow *structure/response correlations* with the aim of eliminating the bias provided by the interdependence of common molecular descriptors. Moreover, the interpretation of regression models should be facilitated as the information encoded in each descriptor is unique.

The similarity/diversity analysis based on previously orthogonalized descriptors is usually called **orthosimilarity** [Randić, 1996b].

📖 [Randić, 1991a, 1991g, 1993b, 1994a; Randić and Seybold, 1993; Pogliani, 1994a, 1994c, 1995b; Randić and Trinajstić, 1994; Amić, Davidović-Amić *et al.*, 1995b, 1997; Lučić, Nikolić *et al.*, 1995a, 1995b, 1995c; Araujo and Morales, 1996a, 1996b, 1998; Šoškić, Plavšić *et al.*, 1996a, 1996b; Mracec, Muresan *et al.*, 1997; Nikolić and Trinajstić, 1998; Ivanciuc, Taraviras *et al.*, 2000; Ivanciuc, Ivanciuc *et al.*, 2000c; González Díaz, Marrero *et al.*, 2003; Fernández *et al.*, 2004; Fernández, Duchowicz *et al.*, 2004; Du, Liang *et al.*, 2005]

- **orthogonalized path counts** \rightarrow orthogonalized descriptors
- **orthogonal Wiener operator** \rightarrow Wiener-type indices
- **orthosimilarity** \rightarrow orthogonalized descriptors
- **Ostwald solubility coefficient** \rightarrow physico-chemical properties (\odot partition coefficients)
- **outdegree** \rightarrow graph
- **ovality index** \rightarrow shape descriptors
- **overall accuracy** \equiv *nonerror rate* \rightarrow classification parameters
- **overall connectivity index** \rightarrow molecular complexity (\odot Bonchev topological complexity indices)
- **overall degree of clustering of a graph** \rightarrow adjacency matrix
- **overall electronic constants σ_m and σ_p** \rightarrow electronic substituent constants
- **overall topological indices** \equiv *Bonchev topological complexity indices* \rightarrow molecular complexity
- **overlap surface** \equiv *common overlap surface* \rightarrow molecular shape analysis (\odot common overlap steric volume)
- **overall Wiener indices** \rightarrow molecular complexity (\odot Bonchev topological complexity indices)
- **overall Zagreb indices** \rightarrow molecular complexity (\odot Bonchev topological complexity indices)