"BABEŞ-BOLYAI" UNIVERSITY CLUJ-NAPOCA
FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING

Ph.D. THESIS

SUMMARY

ATENA PÎRVAN (CĂS. MOLDOVAN)

SCIENTIFIC COORDINATOR
Prof. Dr. MIRCEA V. DIUDEA

2017

"BABEŞ-BOLYAI" UNIVERSITY CLUJ-NAPOCA
FACULTY OF CHEMISTRY AND CHEMICAL ENGINEERING

TOPOLOGY OF THE NANOWORLD

SUMMARY

"Imagination is more important than knowledge.
For knowledge is limited to all we now know and understand,
while imagination embraces the entire world,
and all there ever will be to know and understand."
Albert Einstein

SCIENTIFIC COORDINATOR
Prof. Dr. MIRCEA V. DIUDEA

2017

CONTENTS

Thesis Abstract

TOPOLOGY OF THE NANOWORLD is an attempt of introducing the complexity of the material nature seen at nano-level (10^{-9} m). Complex structures are designed by map operations starting from Platonic solids. Topological analysis was performed by topological symmetry tools, vertex centrality counting, vertex ring signature, RSI, and figure count, in agreement with the Euler's theorem. The main results are presented in two chapters of personal contribution, *IV. Structural Chemistry* and *V. QSAR studies.* The study resulted in 7 published articles, a chapter in a Springer book, 3 sent articles, 6 oral conference presentations, 3 conference posters.

Keywords: graph, multishell-cluster, vertex equivalence class, topological index, layer-matrix, centrality index, topological index, hypermolecule, TOPO GROUP CLUJ algorithm, cluster similarity prediction.

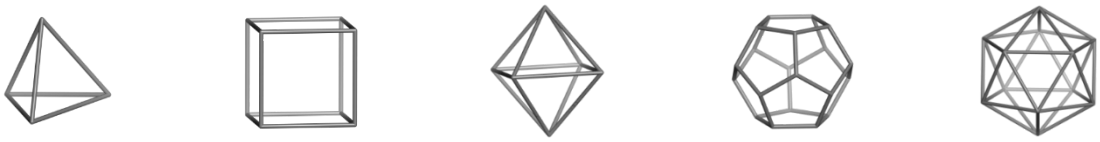
I. Chemical Graph Theory

Chemical Graph Theory is an interdisciplinary science, between Chemistry and Mathematics, with tools from Graph Theory, Set Theory and Statistics, in an attempt to solve difficult problems of Chemistry, like isomer enumeration, symmetry, and finally structure elucidation. Energetics of molecular or ionic structures, their stability and reactivity make the subject of Quantum Chemistry and is out of the aim of this thesis. For a better understanding of this work, some basic definitions in Graph Theory could be useful.

I.7. Polyhedra and Polytopes

A regular polyhedron is a highly symmetric structure, being vertex-transitive, edge-transitive and face-transitive (Coxeter, 1973). There are three symmetry groups: *tetrahedral*; *octahedral* (or cubic) and *icosahedral* (or dodecahedral).

There are five regular polyhedra, known as *Platonic solids* (Figure I.11): tetrahedron, cube, octahedron, dodecahedron and icosahedron can be written as {3,3}; {4,3}; {3,4}; {5,3} and {3,5} by using the basic (Schläfli,1901)



Tetrahedron	Cube	Octahedron	Dodecahedron	Icosahedron
T 3^3 T_d	C 4^3 O_h	O 3^4 O_h	D 5^3 I_h	I 3^5 I_h

Fig. I.11. The Platonic solids, with symbols, vertex configurations and point group symmetry

Archimedean solids (Table I.4 and Figure I.12) are highly symmetric, semi-regular convex polyhedra, with two (or more) types of regular polygons meeting in identical vertices; they are vertex-transitive but not face-transitive.

Table I.4. Archimedean solids

No	Symbol	Polyhedron	Obtaining formula
1	TT	Truncated tetrahedron	$t(T)$
2	TO	Truncated octahedron	$t(O) = t(m(T))$
3	TC	Truncated cube	$t(C) = t(d(m(T)))$
4	TI	Truncated icosahedron	$t(I) = t(d(p_5(T)))$
5	TD	Truncated dodecahedron	$t(D) = t(p_5(T))$
6	CO	Cuboctahedron	$m(C) = m(O) = m(m(T))$
7	ID	Icosidodecahedron	$m(I) = m(D) = m(s(T))$
8	RCO	Rhombicuboctahedron	$m(CO) = m(m(C)) = d(p_4(C))$
9	RID	Rhombicosidodecahedron	$m(ID) = m(m(I)) = d(p_4(I))$
10	TCO	Truncated cuboctahedron	$t(CO) = t(m(m(T)))$
11	TID	Truncated icosidodecahedron	$t(ID) = t(m(s(T)))$
12	SC	Snub cube	$s(C) = d(p_5(C))$
13	SD	Snub dodecahedron	$s(D) = d(p_5(D))$

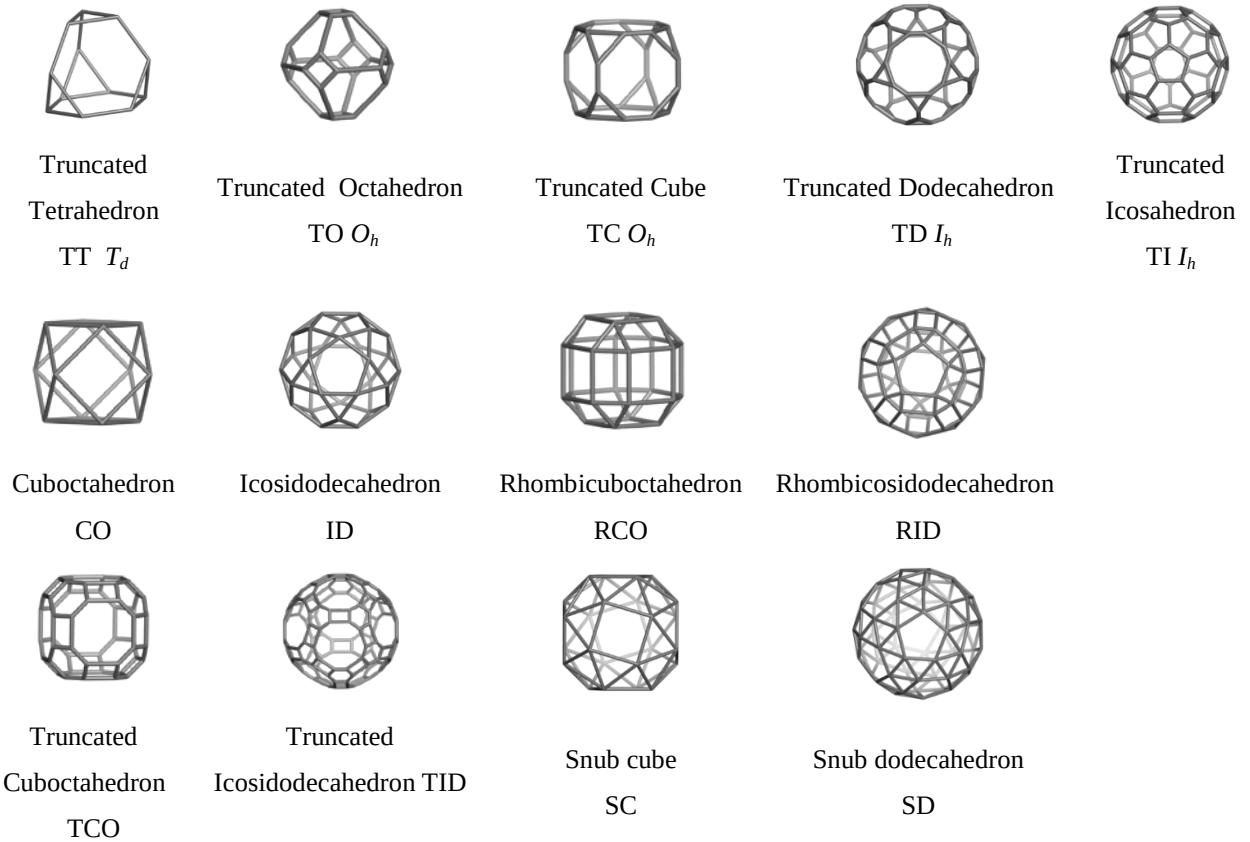


Fig. I.12. Archimedean solids

Catalan solids **(**Catalan, 1865) (Figure I.13) are the duals of Archimedean solids. The Catalan solids are all face-transitive but not vertex-transitive.

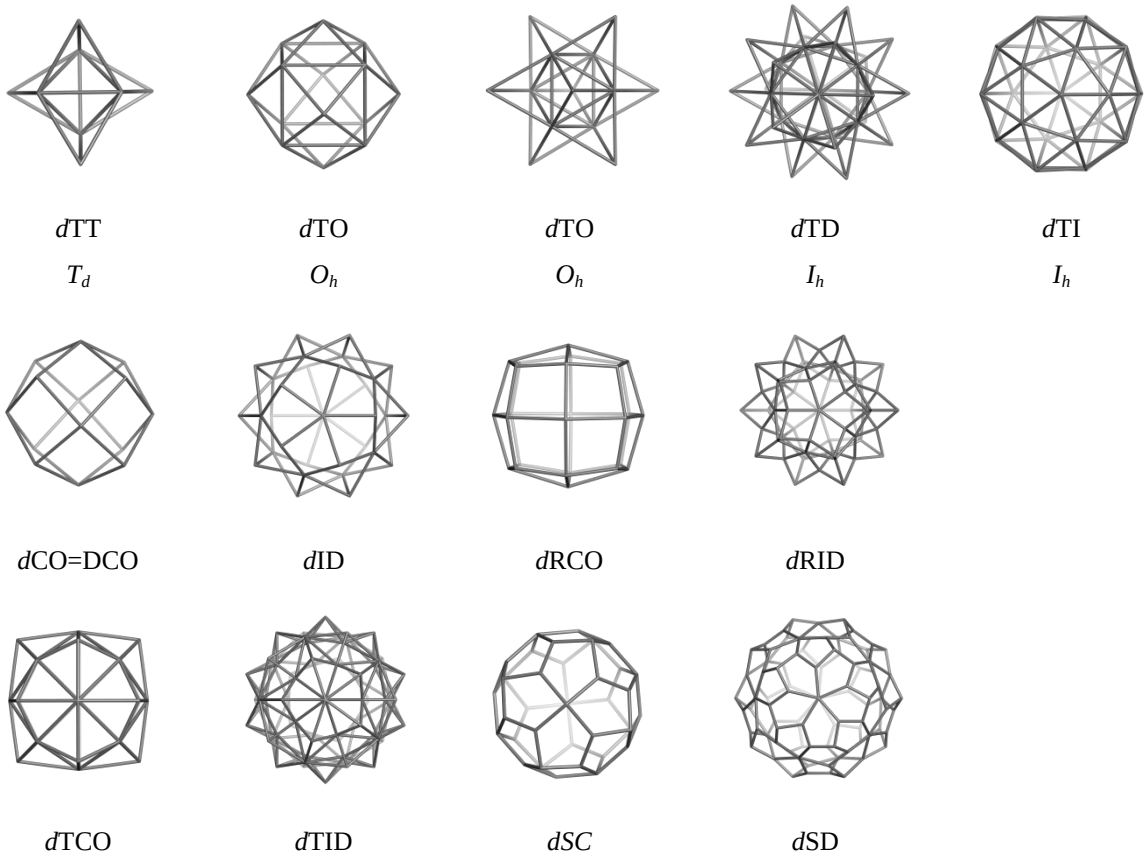


Fig. I.13. Catalan solids

Generalization of a polyhedron to *n*-dimensions is called a polytope (Grünbaum, 2003; Coxeter, 1973). A regular 4-polytope can be written as {α, β, γ}, meaning: γ -polyhedra (of {α,β}) type meet at any edge of the polytope. There are six regular 4-polytopes: 5-Cell {3, 3, 3}, 8-cell {4, 3, 3}, 16-cell {3, 3, 4}, 24-cell {3, 4, 3}, 120-cell {5, 3, 3} and 600-cell {3, 3, 5}. Besides 24-cell, all can be associated to the Platonics. 5-Cell {3, 3, 3} and 24-cell {3, 4, 3} are self-dual.The others (8-cell & 16-cell), (120-cell & 600-cell) are pairs.

There are three types of convex regular polytopes in dimensions 5 and higher, as follows.

The ***n-simplex*** (Coxeter 1973), with the Schläfli symbol $\{3^{n-1}\}$, and the number of k-faces $\binom{n+1}{k+1}$, is a generalization of the triangle or tetrahedron to n-dimensions. A simplex may be defined as the smallest convex set containing the given vertices.

The ***hypercube***, also called an n-cube and denoted Q_n, is a regular polytope with mutually perpendicular sides; it has the Schläfli symbol $\{4,3^{n-2}\}$ and the number of k-faces given by $2^{n-k}\binom{n}{k}$. After Cartesian product graph of n edges: $(K_2)^{\square n} = Q_n$, one obtains the hypercube.

The ***n-orthoplex*** or cross-polytope **(**Coxeter, 1973**)** has the Schläfli symbol $\{3^{n-2},4\}$ and the number of k-faces $2^{k+1}\binom{n}{k+1}$; it exists in any number of dimensions and is the dual of n-cube.

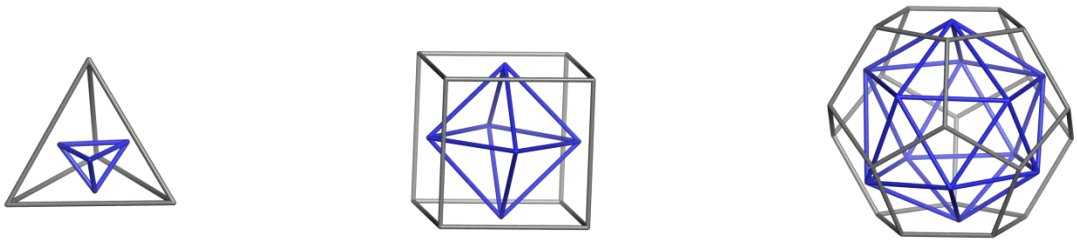
An *abstract polytope* is a structure which considers only the combinatorial properties of a traditional polytope: properties like angles, edge lengths, etc. are disregarded. No n-dimensional space, such as Euclidean space, is required within this theory, in which combinatorial properties are expressed as partially ordered sets or "posets". The poset theory originates in the PhD thesis of Schulte (1985, 2014) and next developed by McMullen and Schulte (2002).

II. Operations on Maps

A map M is a discretized surface domain; several operations on maps are used to modify the topology of a parent map. By running such operations the symmetry of parents is preserved.

II.1 Dual *d*

Dual: put a point in the center of each face of a map/polyhedron, join two such points if their corresponding faces share a common edge. It is the Poincaré *dual* $d(M)$;



dT = T dC = O dD = I

Fig. II.1 Platonic solids as dual-pairs

II.2 Medial *m*

Medial: mark the midpoints of parent edges and join two such points if the edges span an angle while the parent vertices are cut off. Medial is a 4-connected graph, symmetric between the parent and its dual, that is $mdM=mM$, (Figure II.2). The figure sequence of transformed map is: $\{e, 2e, e+2\}$, with e being the number of edges in the parent map.

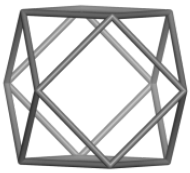
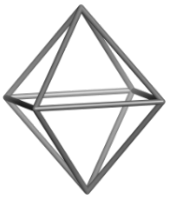
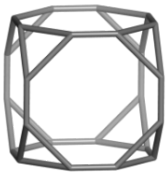
		
C	CO = *m*C = *md*C = *m*O; $(3.4)^2$	O

Fig. II.2 Medial *m* operation (*ambo*; *rectification*) provides a graph symmetric between a polyhedron and its dual.

II.3 Truncation *t*

Truncation: cut off the neighborhood of each vertex by a plane close to the vertex, such that it intersects each edge incident in the point. The resulted truncated polyhedron is always a three-connected one (Figure II.3); its figure sequence is: $\{2e, 3e, e+2\}$.



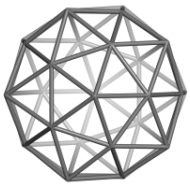
t(C); 3.8^2

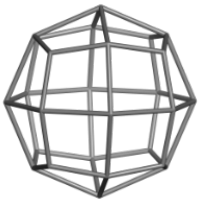
t(O); 4.6^2

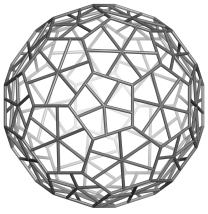
Fig. II.3 Truncation of Cube (left) and Octahedron (right)

II.4 Polygonal mapping p_n

Add a new point in the center of each face. Put *n*-3 points on the boundary edges. Connect the central point with one vertex on each edge (the endpoints included. Figure II.4 illustrates examples of the p_n operations (Diudea and Nagy, 2007).





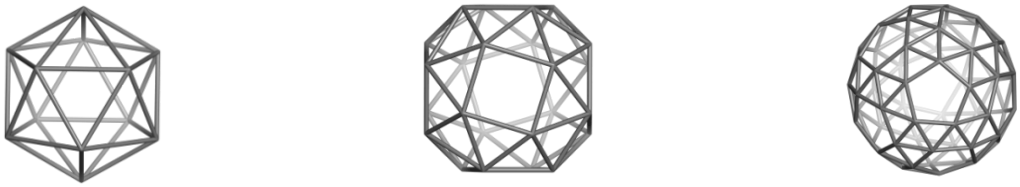


p_3D p_4C= p_4O p_5C_{60}

Fig. II.4 Polygonal p_n operations on Dodecahedron

II.5 Snub *s*

Snub operation is the dual of p_5 operation: $sM = dp_5M$. Similar to the medial operation, $sM = sdM$. (Figure II.5).

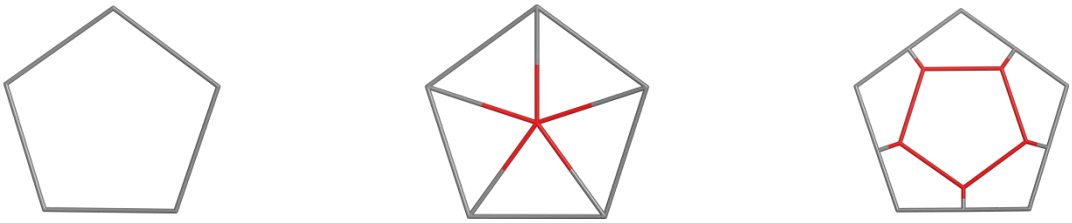


*s*T=I *s*C= *s*O *s*D= *s*I

Fig. II.5 Snub of Platonic solids

II.6 Leapfrog *l*

Leapfrog (or *tripling* or also *dual of kiss*) is a mixed operation (Eberhart, 1891; Fowler, 1986): $lM = d(stM) = t(dM)$. (Figure II.6.)



f_5 $st(f_5)$ $d(st(f_5))$

Fig. II.6 The leapfrog *l* operation on a pentagonal face f_5.

II.7 Quadrupling *q*

Quadrupling (Eberhart, 1891; Diudea and John, 2001**),** also named *chamfering c* (Conway notation) can be written as: $qM = t_{sel}(p_3M)$, where t_{sel} is the selective truncation of the central point added by stellation); (Figure II.7).

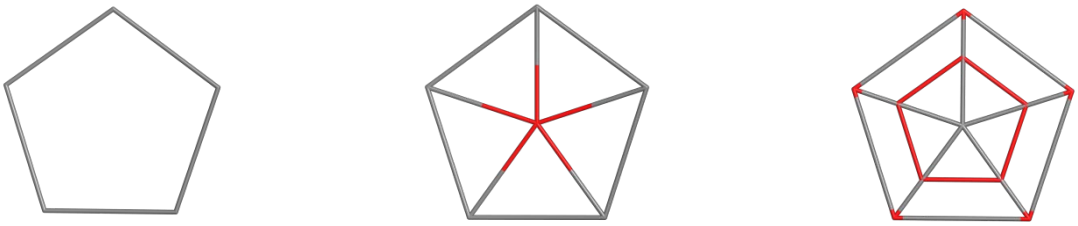


Fig. II.7. The Quadrupling *q* operation on a pentagonal face f_5.

II.8 Septupling

Two septupling operations are known; they multiply seven times the vertices of a three-connected parent map: s_1 and s_2.

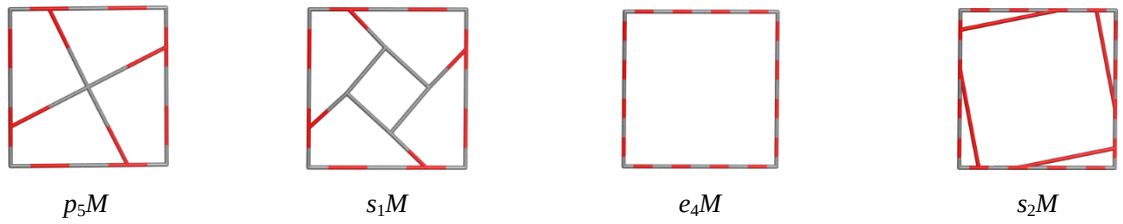


Fig. II.8. Sequence of transforming

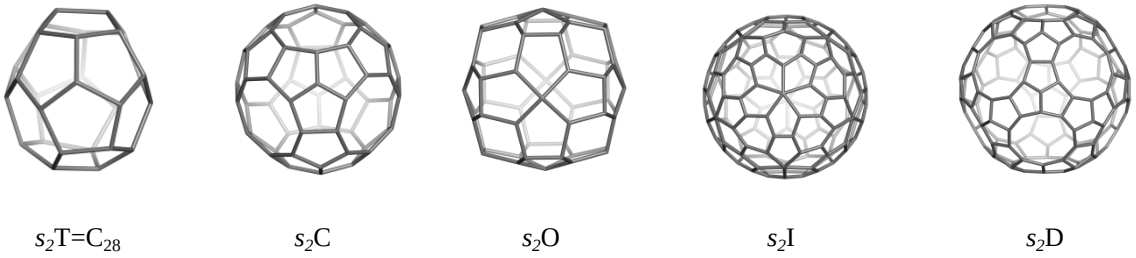
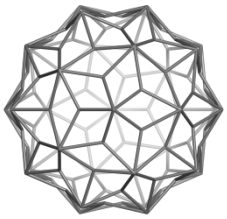


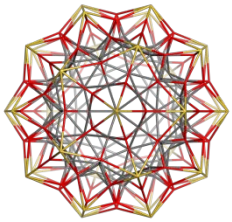
Fig. II.9 Septupling of Platonic solids

II.9. Propelling

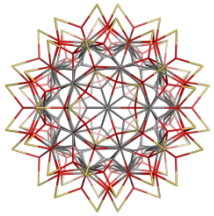
Join by a new point the vertices lying opposite diagonal in a rhomb of a Rh-cage to get Prp1 generation (possible A and B isomers, as there are two diagonals) (Diudea *et al*, 2017). In a second step, put a new point opposite to a vertex of degree higher than 2 and join the new point with the vertices of $d = 2$ surrounding that vertex of $d > 2$, (see Figure II.10).



d(*m*C_{60}).92



ppl(*d*(*m*C_{60}).92).214



ppl(*ppl*(*d*(*m*C_{60}).92).214).242

Fig. II.10. Exemples of structures obtained by propelling

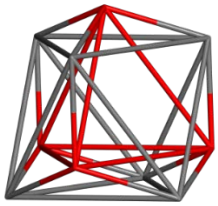
Several software programs were developed by TOPO GROUP CLUJ specialized in polyhedral tessellation and embedment in surfaces of various types, either as finite or infinite structures: TORUS, CageVersatile_CVNET, JSCHEM, Omega Polynomial Counter and NANO-Studio (Diudea *et al*, 2003; Stefu and Diudea, 2005; Nagy and Diudea, 2005; Cigher and Diudea, 2006; Nagy and Diudea, 2009).

Personal Contributions

IV. Structural Chemistry

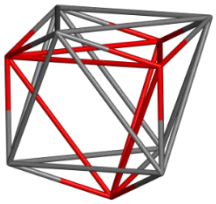
IV.1 Cell@cell higher dimensional structures

Point-centered clusters represent cage-duals of polyhedra with the same number of cells around a central one; they are objects of Euclidean 4D-space (Fathalikhani *et al*, 2016 (Tables IV.1-8) using Euler-Poincare formula *v-e+f=2(1-g)*.This idea can be extended to cages other than Platonics (Figure IV.1-8).



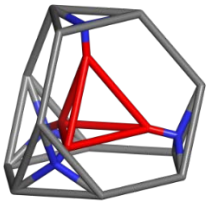
d(TP).10

T@O.10



m(TP).10

T@O.10



t(TP).20

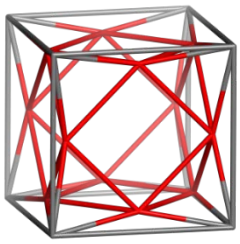
T@((TT)$_4$;(T)$_4$)@TT.20

Figure IV.2. TP-derived structures

Table IV.2. Figure count for clusters derived from TP cluster

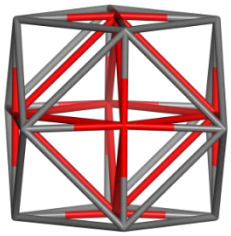
T-structure	v	e	f_3	f_5	f_6	f	p_1	p_2	p_3	M	c	χ	Sym	*pk*; (M)
TP.5	5	10	10	0	0	10	4	0	0	1	5	0	3	T; 0; 0 (P; T)
d(TP).10	10	30	30	0	0	30	4	4	0	2	10	0	3	T; Oct=AP$_3$; 0 (T;O)
m(TP).10	10	30	30	0	0	30	4	4	0	2	10	0	3	T; Oct=Ap$_3$; 0 (T;O)

t(TP).20	20	40	20	0	10	30	4	4	0	2	10	0	3	T;TT; 0 (T; TT)



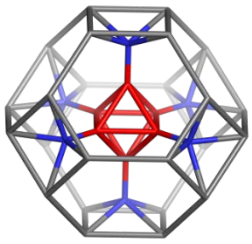
d(OP).20

CO@((Ap4)$_6$;(T)$_8$)@C.20



m(OP).18

O@((O)$_8$;(Py4)$_6$)@CO.18



t(OP).36

O@((TT)$_8$;(Py4)$_6$)@TO.36

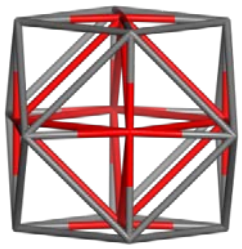
Fig. IV.3 . OP-derived structures

Table IV.3. Figure count for clusters derived from OP cluster

OP-structure	v	e	f_3	f_5	f_6	f	p_1	p_2	p_3	M	c	χ	Sym	*k*; (M)
OP.7	7	18	20	0	0	20	8	0	0	1	9	0	4	T;0;0 (P;O)
d(OP).20	20	60	44	12	0	56	8	6	0	2	16	0	4	T;Ap4;0 (CO;C)
m(OP).18	18	60	52	6	0	58	8	6	0	2	16	0	4	O; Py4 (Oct; CO)
t(OP).36	36	78	32	6	20	58	8	6	0	2	16	0	4	TT;Py4;0 (O;TO)

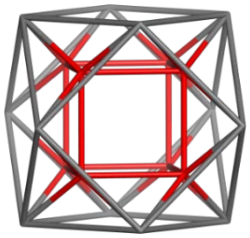
Table IV.4. Figure count for clusters derived from CP cluster

C-structure	v	e	f_3	f_5	f_6	f	p_1	p_2	p_3	M	c	χ	Sym	*pk*; (M)
CP.9	9	20	12	6	0	18	0	6	0	1	7	0	4	0;Py4; 0 (P;C)
d(CP).18	18	60	52	6	0	58	8	6	0	2	16	0	4	Oct; Py4 (Oct; CO)
m(CP).20	20	60	44	12	0	56	8	6	0	2	16	0	4	T;Ap4;0 (C; CO)
t(CP).40	40	80	32	6+6	12	56	8	6	0	2	16	0	4	T;HTO;0 (C; TC)



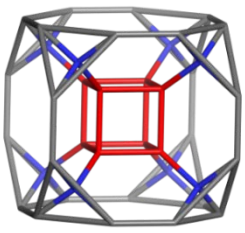
d(CP).18

O@((O)8;(Py4)$_6$)@CO.18



m(CP).20

C@((Ap4)$_6$;(T)8)@CO.20



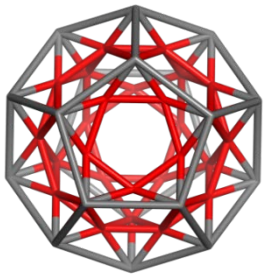
t(CP).40

C@((HTO)$_6$;(T)$_8$)@TC.40

Figure IV.4.CP-derived structures

Table IV.5. Figure count for clusters derived from IP cluster

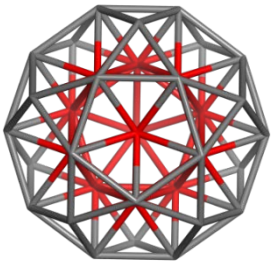
IP-structure	v	e	f_3	f_5	f_6	f	p_1	p_2	p_3	M	c	χ	Sym	p_k; (M)
IP.13	13	42	50	0	0	50	20	0	0	1	21	0	5	T; 0; 0 (P; I)
d(IP).50	50	150	110	24	0	134	20	12	0	2	34	0	5	T;AP5;0 (ID.30;D)
m(IP).42	42	150	130	60	12	142	20	12	0	2	34	0	5	Oct; Py5;0 (I;ID30)
t(IP).84	84	192	80	0	62	142	20	12	0	2	34	0	5	TT; Py5; 0 (I;C_{60})



d(IP).50

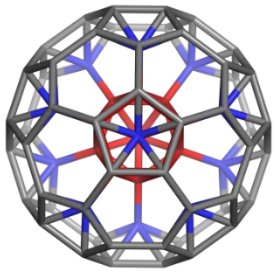
ID@((Ap5)$_{12}$;(T)$_{20}$)@D.50

m(DP).50=D@((Ap5)$_{12}$;(T)$_{20}$)@ID.50



m(IP).42

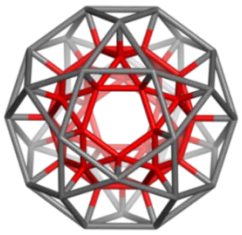
I@((O)$_{20}$;(Py5)$_{12}$)@ID.42



t(IP).84

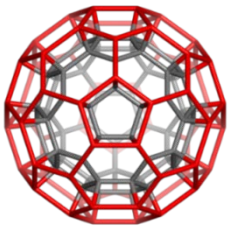
I@((TT)$_{20}$;(Py5)$_{12}$)@C_{60}.84

Fig IV.5. IP-derived structures



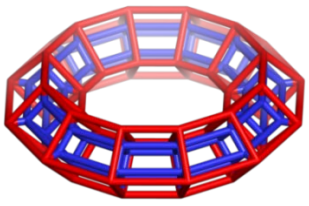
ID@ID.60

ID@((P5)$_{12}$;(P3)$_{20}$)@ID.60

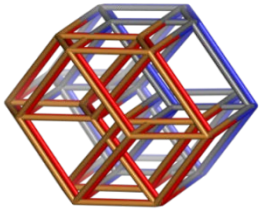


C_{60}@C_{60}.120

C_{60}@((P5)$_{12}$;(P6)$_{20}$)@C_{60}.120



T(4,12)Q4T5.96



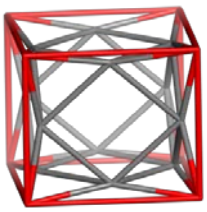
Q6(TU(4,8)Q6T7).64

Fig. IV.7 . Other cell-in-cell structures

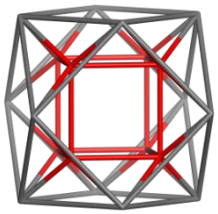
Table IV.7 Figure count for the objects in Figure IV.7.

Structure	0	1	f3	f4	f5	f6	2	p3	p5	p6	m	3	4	5	χ
ID@ID.60	60	150	40	60	24	0	124	20	12	0	2	34	0	0	0
C_{60}@C_{60}.120	120	240	0	90	24	40	154	0	12	20	2	34	0	0	0
T(4,12)Q_4T_5.96	96	240	0	216	0	0	216	0	0	0	0	84	12	0	0*
Q6.64	64	192	0	240	0	0	240	0	0	0	0	160	60	12	0

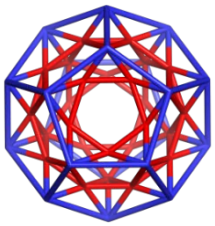
* in case of Torus, the right member of (2) gives all time zero, because the torus is a surface of genus g=1.



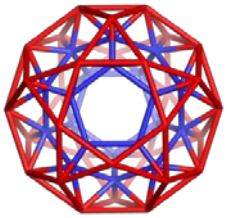
CO@C.20
CO@((AP4)$_6$;(T)$_8$)@C
m(CP).20



C@CO.20
C@((AP4)$_6$;(T)$_8$)@CO
d(OP).20



ID@Do.50
ID@((AP5)$_{12}$;(T)$_{20}$)@Do
m(DP).50



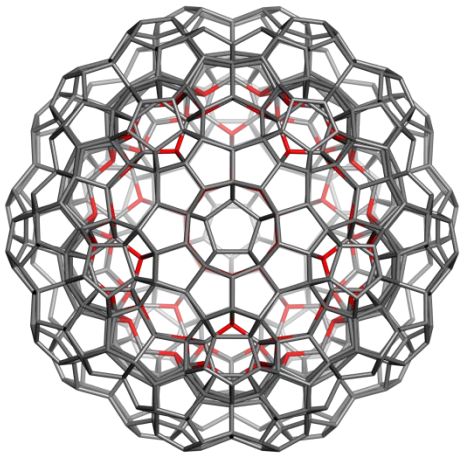
D@ID.50
D@((AP5)$_{12}$;(T)$_{20}$)@ID
d(IP).50

Fig. IV.8. Sphere inversion by 4D-clusters: moving on the fourth dimension

IV.2. Covering in icosahedral clusters

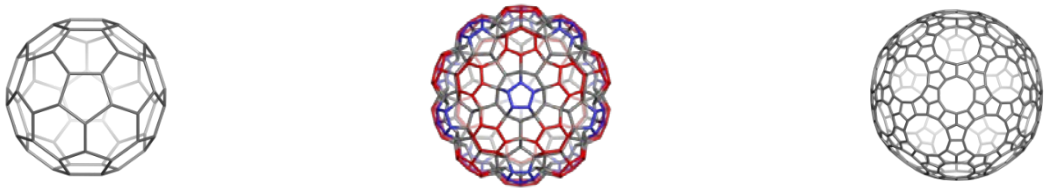
Design of multi-shell cages

The operations sequence in building the cluster C_{750}, achived by map operations is : *ts*(*p*$_4$(C_{60})).330; *s*$_2$(C_{60}).420; *ts*(*p*$_4$(C_{60}))@s$_2$(C_{60}).750 (Figure IV.9). Cluster C_{750} is a spongy one, it is a C_{20} tessellation:



ts(*p*$_4$(C_{60}))@(s$_2$(C_{60}).420).750

$C_{60}((C_{20})_{60})$.750



C_{60} $ts(p_4(C_{60}).330$ $s_2(C_{60}).420$

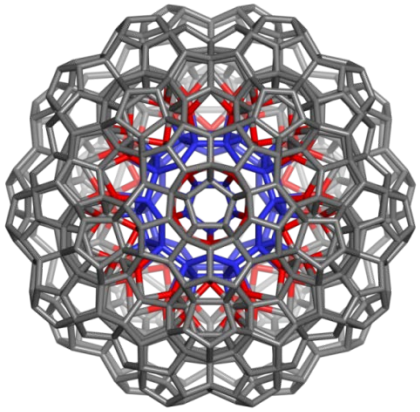
Fig.IV.9. Multi-shell structures on 750 atoms

Table IV.9. Symmetry of C_{750}: Automorphism group = $C_2 \times A_5$ = Ih ; |Ih| = 120.

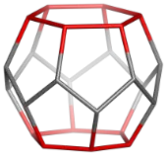
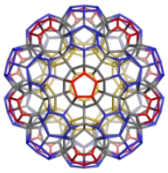
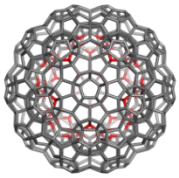
Class	Centrality signature (5;6)	No.Elements	Vertex degree	Atom type
1	0.0425537487829127	60	4	5^5
2	0.0425405656366799	30	4	5^5
3	0.0408741428983785	60	3	5^3
4	0.0403249632533878	60	4	5^6
5	0.0403215210989583	60	4	5^5.6
6	0.0403184110690464	60	4	5^5.6
7	0.0380980964599947	60	4	5^5
8	0.0380776127196794	60	4	5^5
9	0.0380525586272046	60	4	5^5
10	0.0363966020960237	60	3	5^3
11	0.0363899446618803	60	3	5^3
12	0.0363398403991418	120	3	5^3

The structure C_{810} (Figure IV.10) has alike structure with C_{750}, only that it has a C_{60} fullerene inside C_{750}, so C_{810}=C_{60}@C_{750}. When merged to the hollow of C_{750} (i.e. $ts(p_4(C_{60}))330$), C_{60} offer C_{20} and C_{24} cells, as shown in the structure $C_{60}(@((C_{20})_{12};(C_{24})_{20}).390$.

The substructures are given at the bottom of figures.



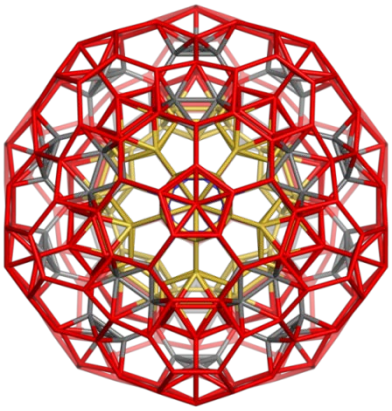
$C_{60}@((C_{20})_{12};(C_{24})_{20})@(C_{20})_{60}.810$ $C_{60}@(C_{60}(C_{20})_{60}).810$

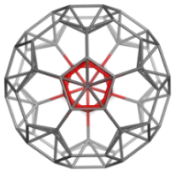
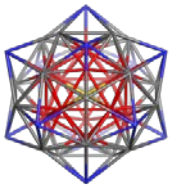
C_{24} $C_{60}(@((C_{20})_{12};(C_{24})_{20}).390$ $C_{60}(D)_{60}750$

Fig. IV.10 C_{810} and its components

A cluster of general formula $M@M_{12}$, namely $C_{408}=C_{84}@(C_{84})_{12}.408=t(Diu45).408$ (Pârvan-Moldovan *et al*, 2014 – Figure IV.11) was obtained by truncation of Diudea's cluster $Diu_{45}=IP@(IP)_{12}.45$

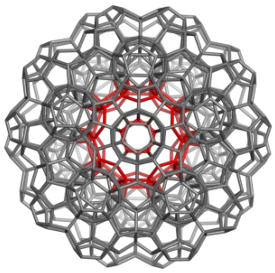


$(I@(TT_{20})@((I)_{12};TT_{20};TT_{30})@(Py5@TT_5)_{12}.408$
$t(Diu_{45}).408$

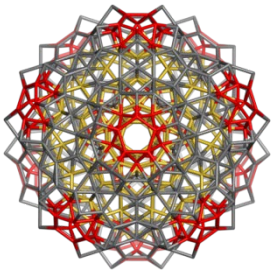




IP.13 t(IP).84 $Diu_{45}=IP@(IP)_{12}.45$

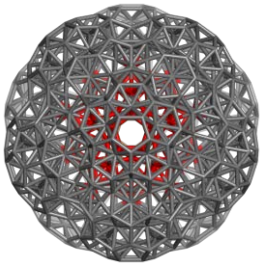
Fig. IV.11 C_{408} and its components



$C_{60}@C_{750}$.810



$C_{130a}@(C_{130a})_{12}$.810
$d(m(\mathrm{Diu}_{45})204).810$

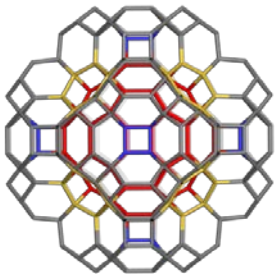


$C_{130a}@(C_{13a})_{12}$.810
$d(t(\mathrm{Diu}_{45})408).810$

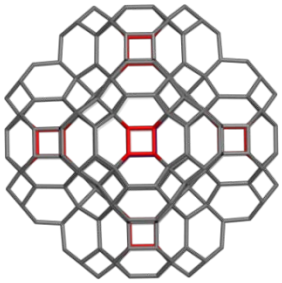
Fig. IV.13: Multi-shell clusters of Icosahedral symmetry C2 x A5 and rank higher than 3

IV. 3. Covering in octahedral clusters

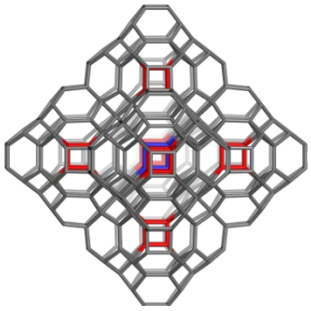
The multi-shell structures in Figure IV.14 have been designed by truncating some Diudea's multi-shell clusters (Figure IV.15), named Diu_k, with *k* being the number of atoms; (Kooperazan-Moftakhar, 2015).



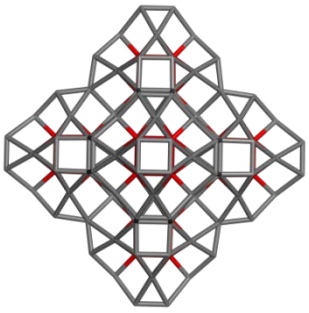
$C_{72}@(C{:}TO_4)_6$.360 spongy
$C_{72}@((C{:}TO_4)_6;(P_6)_8)$.360
$t(m(p_3(C)P^{14}15).50).360$



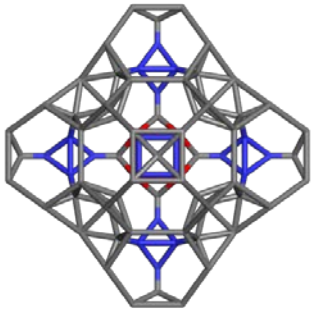
$C_{88}@(C_{72}@(C{:}TO_4)_6.360).376$
$t(\mathrm{Diu}_{51}).376$



$C_{88}@(C_{88})_6$.424
$C@(C_{88})_6$.424
$t(\mathrm{Diu}_{57}).424$

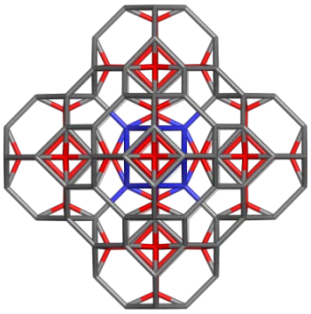


$C_{44}@(C_{44})_6$.212
$C@(C_{44})_6$.212
$m(\mathrm{Diu}_{57}).212$



C@(Oct@(TT)$_8$.36)$_6$.160

t(Diu$_{21}$).160



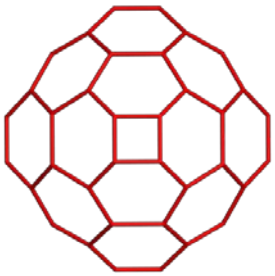
CO@(Oct@(TT)$_8$.36)$_6$.192

t(Diu$_{25}$).192

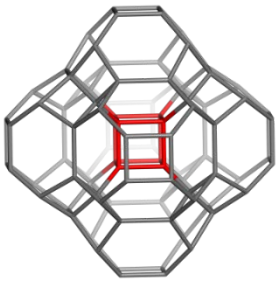
Figure IV.14 . Octahedral multi-shell structures.

The only structure made by Medial operation is C$_{44}$@(C$_{44}$)$_6$.212, also named *m*(Diu$_{57}$).212 (Figure IV.14).

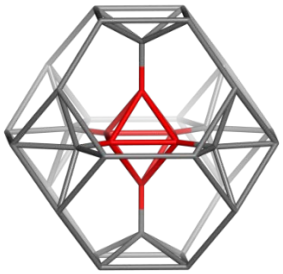
The structure C$_{72}$@(C:TO$_4$)$_6$.360 is a spongy one; its hollow fits precisely the cage C$_{72}$ = (4:6$_4$)$_6$;6$_8$.72 and also C$_{88}$, the last one being named C@TO$_6$.88 (i.e. six truncated octahedra joined at the faces of the cube).



C$_{72}$ = (4:6$_4$)$_6$;6$_8$.72



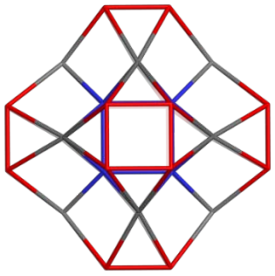
C$_{88}$=C@TO$_6$.88



O@(TT)$_8$.36

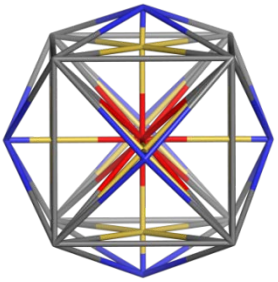
O@((TT)$_8$;(Py$_4$)$_6$)@TO.36

t(OP).36

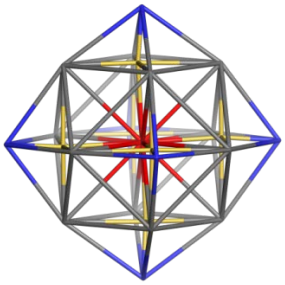


C$_{44}$=C@CO$_6$.44

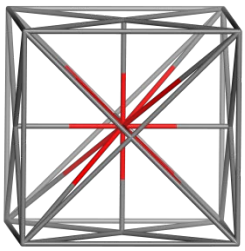
d(O@(O)$_8$.18).52X.44



p_3(p_3 (C)P^{14}15).21=Diu$_{21}$

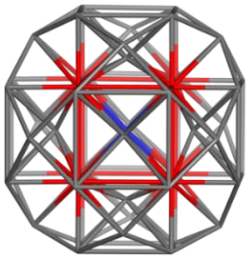


p_3 (p_3 (CO)P^{18}.19).25=Diu$_{25}$



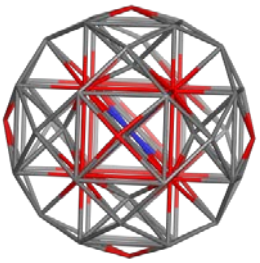
p_3 (C)P^{14}.15

$P^{14}@T_{24}$.15



$m(p_3$ (C)$P^{14}15)P^8$.51= Diu_{51}

p_3 (C)P^8.15@(P^8@$(O)_4$13)$_6$.51

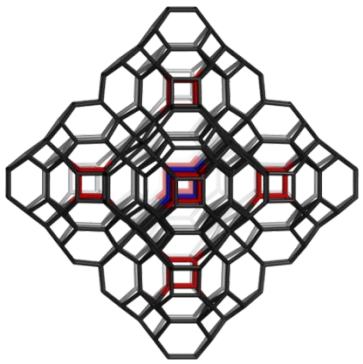


(P^8@$(O)_6$15)@(P^8@$(O)_6$15)$_6$.57

(p_3 (C)P^8.15)@(p_3 (C)P^8.15)$_6$.57

p_3 (Diu_{51})=Diu_{57}

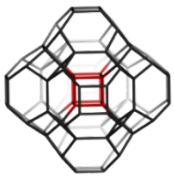
Fig. IV.15. Substructures and multi-shell clusters (Diu_k) of octahedral symmetry.



C_{88}@(C_{88})$_6$.424

C_8

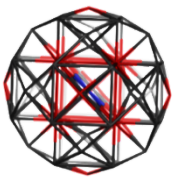
C@(C_{88})$_6$.424



C_{88}=C@(TO)6.88

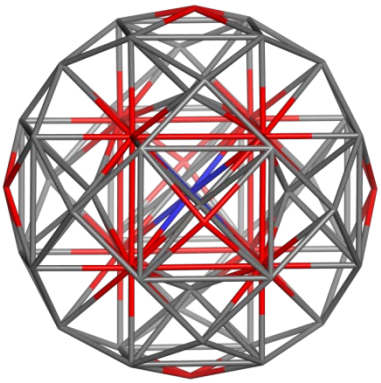
$t(p_3$(C)$P^{\wedge 8}$.15).88

t(Diu_{57}).424



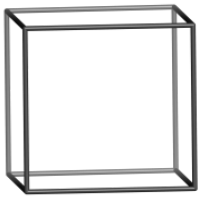
Diu_{57}=(p_3 (C)P^8.15)@(p_3 (C)$P^{\wedge 8}$.15)6.57

Fig. IV.16. Cluster C_{424} and its substructures

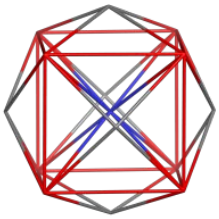


p_3(C)P$^{\wedge 8}$.15)@(p_3 (C)P$^{\wedge 8}$.15)$_6$.57 (P$^{\wedge 8}$@Oct$_6$.15)@(P$^{\wedge 8}$@Oct$_6$.15)$_6$.57 Diu$_{57}$= p_3(Diu$_{51}$)

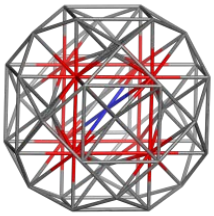
Fig IV.18. Cluster Diu$_{57}$ and its substructures



C$_8$

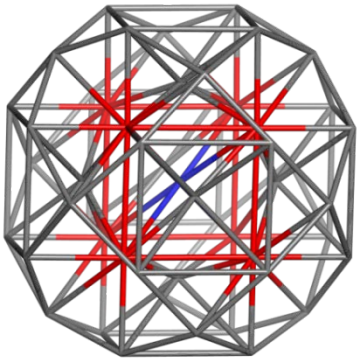


p_3(C)P$^{\wedge 8}$.15
P$^{\wedge 8}$@Oct$_6$.15

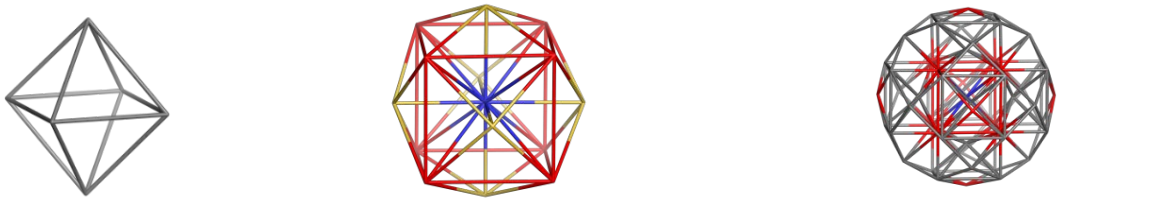


Diu$_{51}$=(P$^{\wedge 8}$@Oct$_6$.15)@
(P$^{\wedge 8}$@Oct$_4$.13)$_6$.51

Fig IV.18. Cluster Diu$_{57}$ and its substructures (continued)



(P$^{\wedge 8}$@(O)$_6$.15)@(P$^{\wedge 8}$@(O)$_4$.13)$_6$.51 m(p_3(C)P$^{\wedge 14}$.15)P$^{\wedge 8}$.51



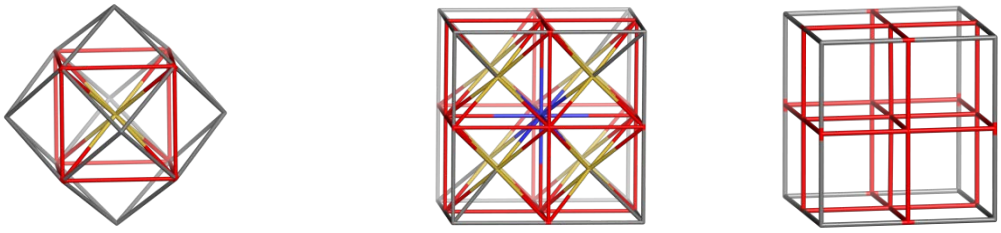
O.6 | p_3(C)P$^{\wedge 14}$.15 P$^{\wedge 14}$@(T)$_{24}$.15 CD(m(O@(O)$_8$.18)60).15 | (P$^{\wedge 8}$@(O)$_6$.15)@(P$^{\wedge 8}$@(O)$_6$.15)$_6$.57

Fig IV.19. Cluster Diu$_{51}$ and its substructures

IV.4. Cube and Rh$_{12}$ networks

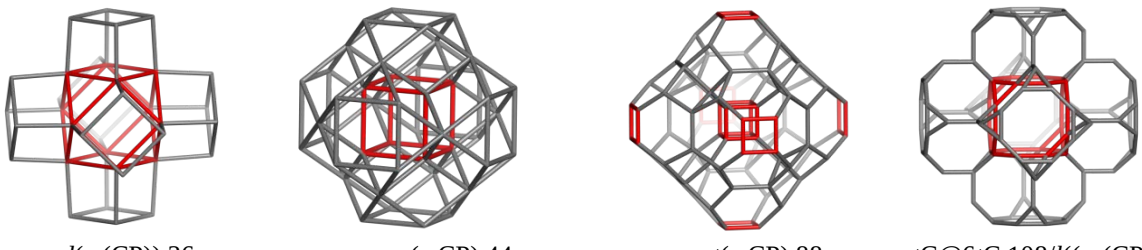
Filling the space by cube and its transforms by symmetry and/or map operations requires consideration of infinite networks and translational methods producing and investigating them.

As a seed, in this section, we take the small cluster P$^{\wedge 8}$@(O)$_6$@p_3C.15, consisting of six octahedral shapes all meeting in the central vertex/point "P" of a capped/starred cube; shortly it is called p_3CP.15 (Pârvan-Moldovan and Diudea, 2015a, b). Its transforms, by dual, medial, truncation and leapfrog represent repeating units of the nets obtained applying these operations, either to the body centered cubic net or to the simple cubic *pcu* net (see CP.222.35 and C.222.27 in Figure IV.20



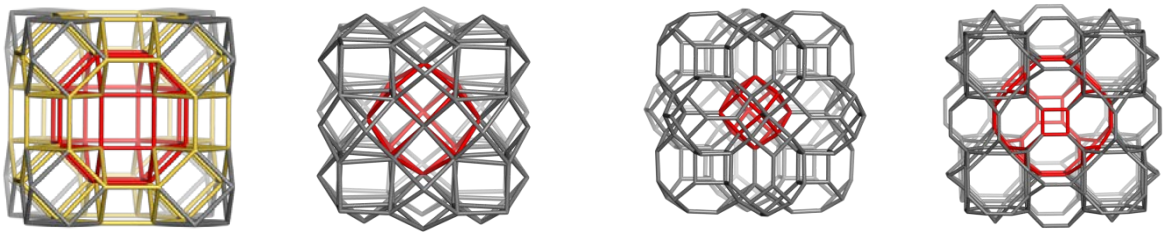
p_3CP.15 | CP.222.35 | *pcu* C.222.27

Fig IV.20. Seed clusters for the cubic net transforming



d(p_3(CP)).36 | m(p_3CP).44 | t(p_3CP).88 | tC@6tC.108/l((p_3(CP)).108

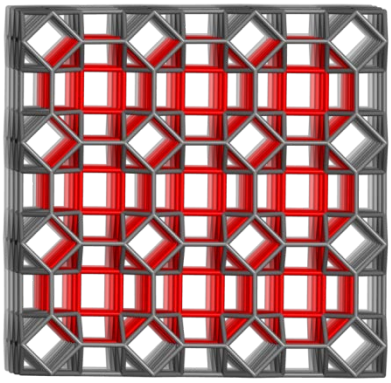
Fig. IV.21. Structural units transformed by map operations.



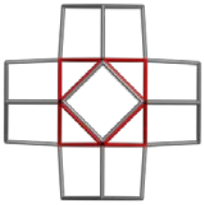
d(P@(CP)$_8$).96 *m*(P@(CP)$_8$).94 *t*(P@(CP)$_8$).172 *l*(P@(CP)$_8$).264

Fig. IV.21. Structural units transformed by map operations (continued)

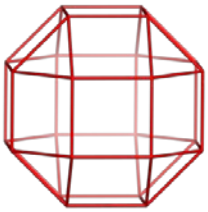
The net/co-net unit pairs are as follows: *dual* net (*d*(p_3(CP)).36 /*d*(P@(CP)$_8$).96; *medial* net (*m*(p_3(CP)).44 /*m*(P@(CP)$_8$).94); *truncate* net (*t*(p_3(CP)).88 /*t*(P@(CP)$_8$).172) and *leapfrog* net (*l*(p_3(CP)).108 /*l*(P@(CP)$_8$).264);



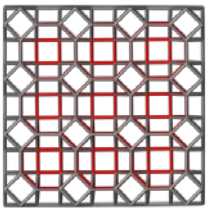
C$_{36}$.444.768.sel.432



d(p_3(CP)).36

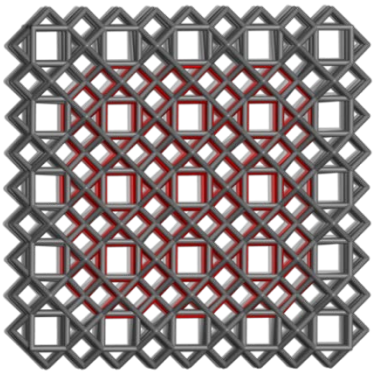


*mm*C.24

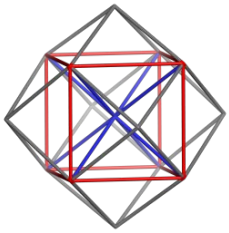


C$_{36}$.444.768.sel.432

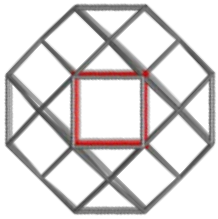
Fig. IV.22. Cubic network transformed by dual



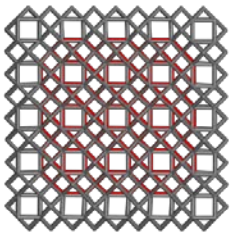
C_{94}.333.1480.sel.576



p_3 (CP).15

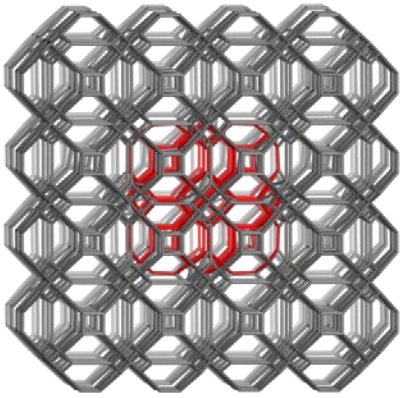


m(*st*CP).44

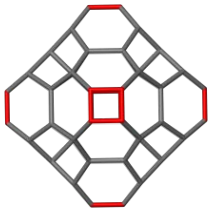


C_{94}.333.1480.sel.576

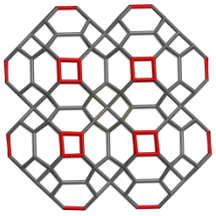
Fig. IV.23. Cubic network transformed by p_3



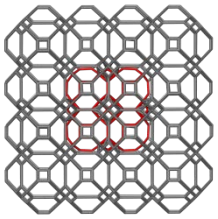
C_{88}.444.2392.sel.428



C_{88}

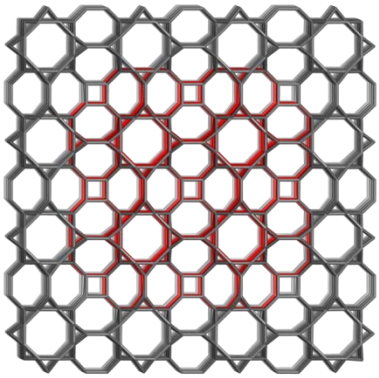


C_{428}

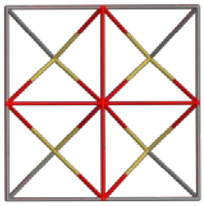


C_{88}.444.2392.sel.428

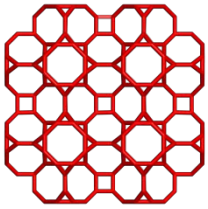
Fig. IV.24. Cubic network transformed by *t*



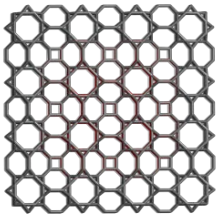
C_{264}.333.2256.sel.576



P@8CP.35



C_{576}



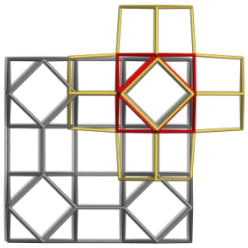
C_{264}.333.2256.sel.576

Fig. IV.25. Cubic network transformed by *l*

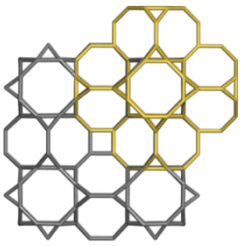
Table IV.18. Figure count in p_3CP.15 derivatives

	v	*e*	3(2)	4(2)	5/6/8(2)	2	1(3)	2(3)	3(3)	M	3	χ	*k*	*n*(3);(M);(4)
1	15	44	36	0	0	36	6	0	0	1	7	0	4	O;(P;*st*C)
2	36	84	20	36	8	64	6	8	0	2	16	0	4	C;hCO;(CO;TO)
3	96	240	64	102	6	172	12	8	6	2	28	0	4	C;CO;h*mm*C;(*mm*C..)
4	44	108	36	36	0	72	6	0	0	2	8	0	4	CO;(C..)
5	94	240	72	96	0	168	12	8	0	2	22	0	4	CO;C;(Rh_{12}..)
6	88	152	0	36	36	72	6	0	0	2	8	0	4	TO;(C..)
7	172	302	0	72	72	144	12	0	0	2	14	0	4	TO;(Rh_{12}..)
8	108	180	44	0	36	80	6	0	0	2	8	0	4	TC;(TC..)
9	264	480	136	6	96	238	20	0	0	2	22	0	4	TC;(*d*(*st*CO).48..)

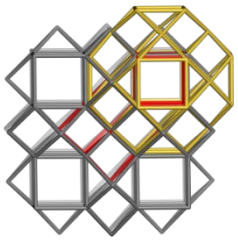
1= p_3CP.15 ; 2= *d*(p_3(CP)).36 ; 3= *d*(P@8CP).96 ; 4= *m*(p_3CP).44 ; 5= *m*(P@8CP).94; 6= *t*(p_3CP).88 ; 7= *t*(P@8CP).172 ; 8= *l*((p_3(CP)).108 ; 9 = *l*(P@8CP).264.



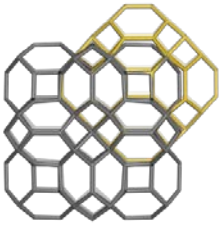
d(p_3(CP)).36/*d*(P@(CP)$_8$).96



l(p_3(CP)).108/*l*(P@(CP)$_8$.264



m(p_3(CP)).44/*m*(P@(CP)$_8$).94

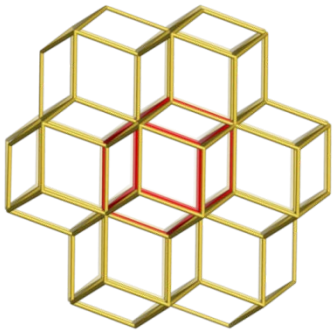


t(p_3(CP)).88/*t*(P@(CP)$_8$).172

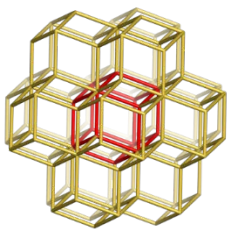
Fig. IV.26. Nets and co-nets for the cubic net relatives

Space filling by Rh_{12} and related structures

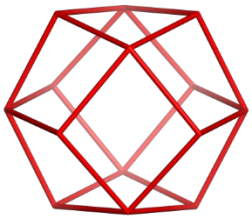
In this section, attention will be focused on the rhombic dodecahedron, Rh_{12}.14, designed as the dual of medial of cube, *d*(*m*C).14 (or dual of cuboctahedron DCO).



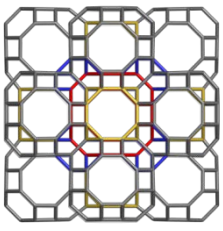
Rh_{12}@12Rh_{12}.94



Rh_{12}@12Rh_{12}.94

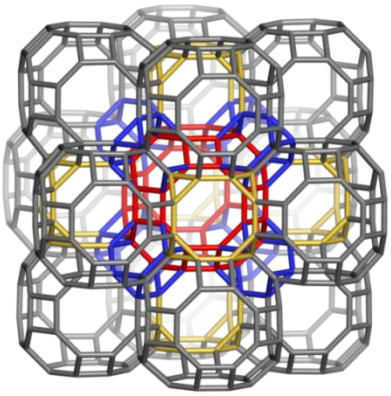


Rh_{12}.14 = *d*(*m*C).14=*d*CO

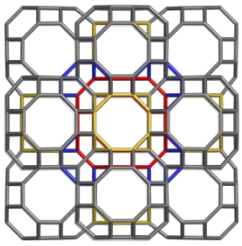


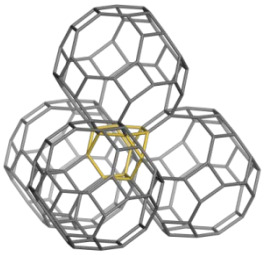
l((Rh_{12}@12Rh_{12}).480

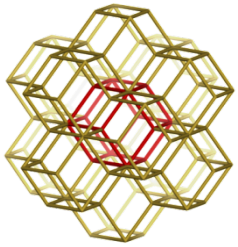
Fig. IV.27. Space filling by cuboctahedron CO relatives



TCO@(6TC;8TT;12TCO).480 | *l*(Rh_{12}@12Rh_{12}).480 | *l*(DCO) = TCO(4); net

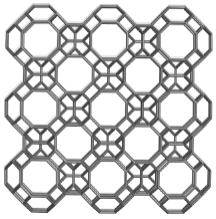


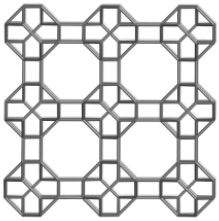




l(Rh_{12}@12Rh_{12}).480.4 | TT@4TCO.168 | Rh_{12}@12Rh_{12}.94

Fig. IV.28. TCO space filling



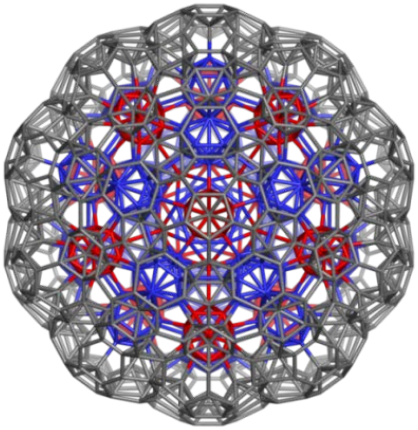


DCO.228 | TCO(4).1248 net | TCO(4).324 co-net

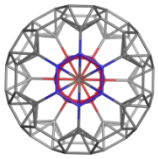
Fig. IV.29. Space filling by cuboctahedron CO relatives

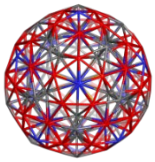
IV.5. C_{60} related multi-shell clusters

An elaborated structure, obtained by truncation of Diu_{125}=p_3(C_{60}@d(C_{60})$P^{\wedge 32}$).125 cluster (Figure IV.30) is shown; it is detailed as C_{1208}=t(Diu_{125}).1208=(d (C_{60}))@(C_{84})$_{12}$;C_{100})$_{20}$).1208.

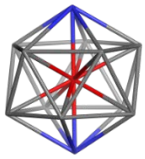


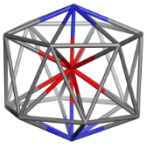
t(Diu$_{125}$).1208 *d*(C$_{60}$)@(C$_{84}$)$_{12}$;C$_{100}$)$_{20}$).1208

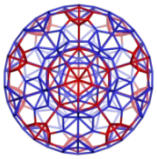




t(*p*$_3$(Ap6)P).100 *t*(IP).84 *p*$_3$(C$_{60}$@(d(C$_{60}$)P$^{\wedge 32}$).125=Diu$_{125}$







IP.13 P@*p*$_3$(Ap$_6$).15 (*d*(C$_{60}$))@((TT)$_{12}$;(T)T$_{20}$)@C$_{180}$(*I*$_h$).244

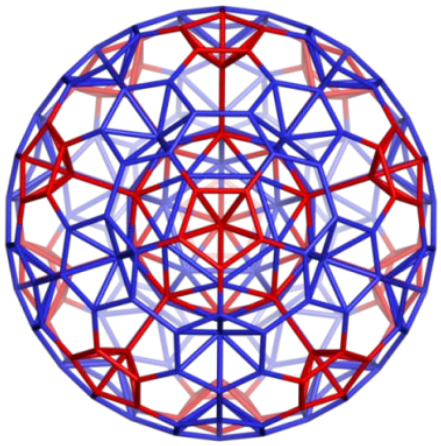
Fig. IV.30. C$_{1208}$ and its components

Table IV.23. Euler formula for some of the studied clusters

Cluster	*v*	*e*	*f*$_3$	*f*$_5$	*f*$_6$	*f*	c1	c2	c3	c4	c5	c	χ	M$_k$;[M]
d(C$_{60}$).32	32	90	60	0	0	60	0	0	0	0	0	0	2	((3^5)$_{12}$;(3^6)$_{20}$)=St(Do)32
C$_{180}$(I$_h$)	180	270	0	12	80	92	0	0	0	0	0	0	2	(5.6^2)60;(6^3)120
C$_{540}$(I$_h$)	540	810	0	12	260	272	0	0	0	0	0	0	2	(5.6^2)60;(6^3)480
P@*p*$_3$(Ap$_5$).13	13	42	50	0	0	50	20	0	0	0	1	21	0	T;0;0;0;[St(Ap5)=Ico.12]
C$_{84}$	84	192	80	12	50	142	20	12	0	0	2	34	0	TT; Py5;0;0;[Ico; C$_{60}$(I$_h$)]
P@*p*$_3$(Ap$_6$).15	15	50	60	0	0	60	24	0	0	0	1	25	0	T;0;0;0;(*p*$_3$(Ap6))
C$_{100}$	100	230	96	12	62	170	24	12	2	0	2	40	0	TT; Py$_5$;Py$_6$;0; [*p*$_3$(Ap$_6$);C$_{72}$(D$_{6d}$)]
C$_{33}$	33	122	150	0	0	150	60	0	0	0	1	61	0	T;0;0; 0;(*p*$_3$ (D)32)
Tr(C$_{33}$).244	244	572	240	12	170	422	60	12	20	0	2	94	0	TT;Py$_5$;Py$_6$;0; [(*d*(C$_{60}$),32;C$_{180}$]
C$_{125}$	125	604	870	0	0	870	390	0	0	0	1	391	0	T;0;0;0; (*p*$_3$(C$_{60}$)92)
C$_{1208}$	1208	3214	1560	12	950	2522	390	24	40	60	2	516	0	TT; (Py$_5$;St(Ap$_5$);

(Py_6; $p_3(Ap_6)$; J52*;
$[d(C_{60})32;C_{540}]$

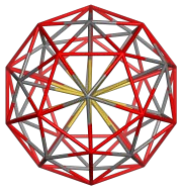
* Johnson object J_{52}; Py_k=Pyramid on k- basis; Ap_k=Antiprism on k-basis.



$(d(C_{60}))@((TT)_{12};(T)T_{20})@C_{180}(I_h).244$ $C_{244}= t(Diu_{33}).244$

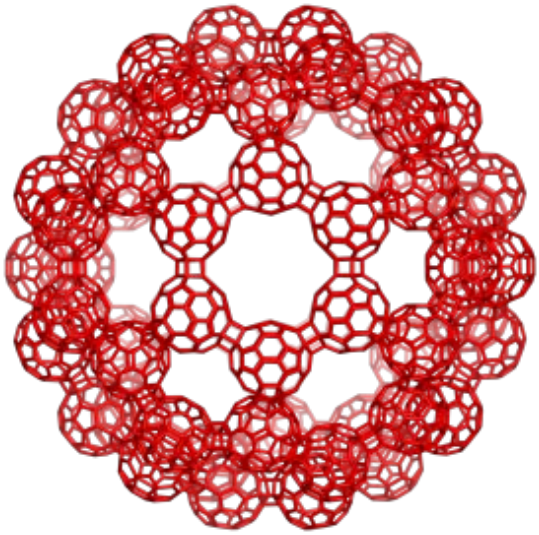
TT.12

D.20

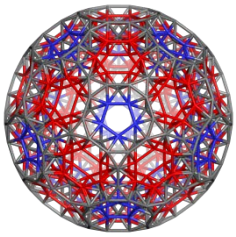


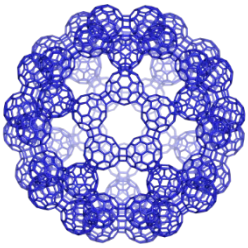
p_3 $(DP^{\wedge 32}).33=Diu_{33}$

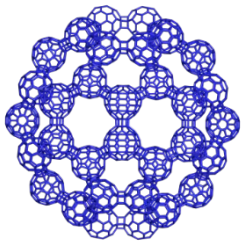
Fig.IV.31. C_{244} and its components



$C_{60}Y(60C_{60}\_P_5).3600$ $C_{60}Y(60C_{60}; 90P_5).3600$ $t(d(C_{750}).630).3600$







$d(C_{750}).630$ C_{3600} (5) C_{3600} (2)

Fig.IV.32. $C_{3600} = t(d(C_{750}).630).3600$

IV.6. Euler characteristic

Genus *g* is involved in Euler characteristic determination of a closed orientable surface (*g* = the number of tori in a connected sum decomposition of the surface, or the number of handles or holes an object has) by the Poincaré formula

$$v - e + f = \chi = 2(1 - g) \qquad (2)$$

Euler characteristic can be computed for general surfaces as the alternating sum of figures of dimension/ rank *k* by finding a polygonization of the surface.

$$\chi = f_0 - f_1 + f_2 - f_3 + \ldots, \qquad (6)$$

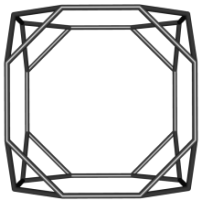
Pairs of map operation

Theorem 1. *Let* $\{v, e, f\}$ *and* $\{n_1e+\delta, n_2e, n_3e\}$ *be types of a parent polyhedron and its derivative polyhedron* $o(P)$ (*obtained by a map operation* o). *Also, let both* P *and* $o(P)$ *have the same Euler characteristic* χ. *Then,* $\delta = \chi$ *if and only if* $(n_1 + n_3) = n_2$. (Pîrvan-Moldovan *et al*, 2016)

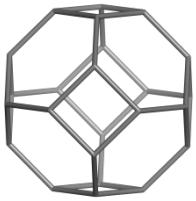
Corolary 1 (2). The dual of the generalized transform *d(o(P))* will have the type: $\{n_3e, n_2e, n_1e+\chi\}$. This comes out from the property of Schläfli symbol that its reversal gives the symbol of the dual polyhedron.

Corolary 1 (3). Difference in the number of vertices of the transformed polyhedral graphs by selected pairs of map operations, o_1 and o_2, equals the Euler characteristic of the embedding surface (Table IV.29-30): $|V(o_1(P))| - |V(o_2(P))|=\chi$.

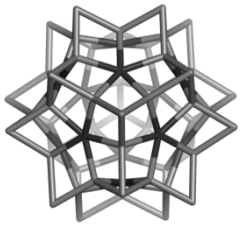
p_4(C).26; (8(4^3).18(4^4))



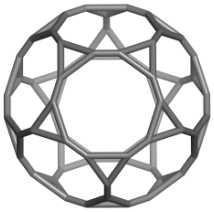
t(C).24; (3.8^2)



l(C).24; (4.6^2)



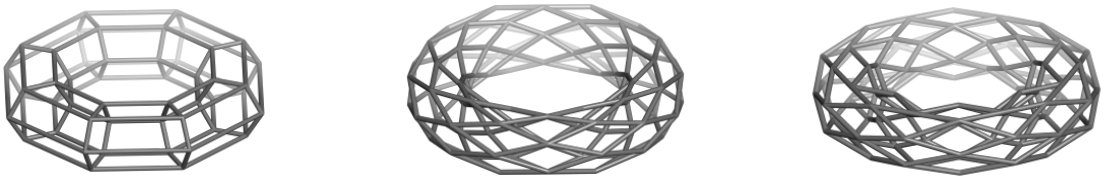
p_4(D).62
(12(4^5).30(4^4).20(4^3))



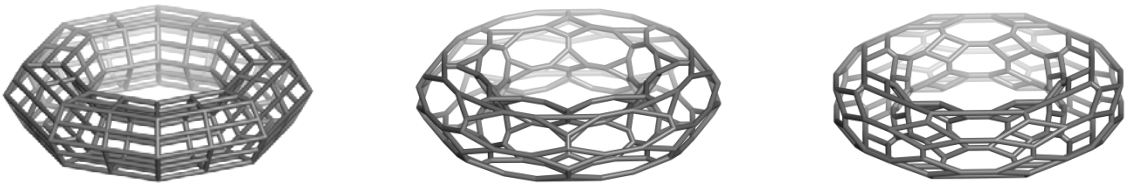
t(D).60 [3.10.10]

l(D).60=C_{60} [5.6.6]

Fig. IV.33. Operations applied on Cube C (top) and Dodecahedron D (bottom): vertex number differece: χ=2 (Euler characteristic of the sphere).

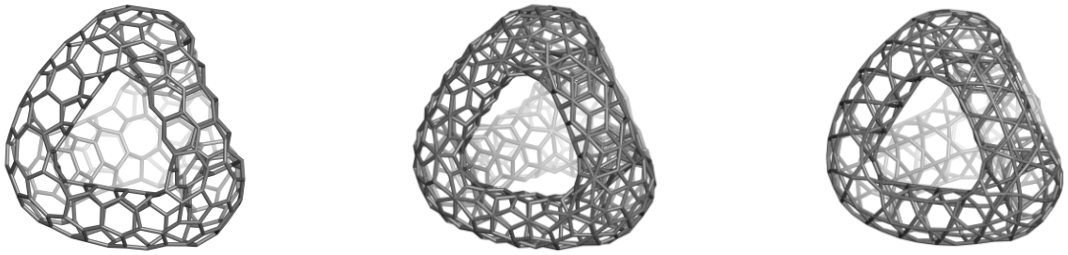


H6.8.48 {4^4}　　*dm*(H6.8).96 {4^4}　　*m*(H6.8).96 {4^4}



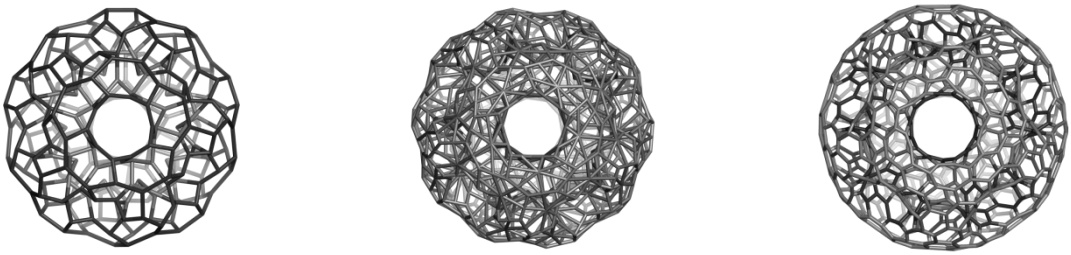
p_4(H6.8.48).192　　*t*(H6.8.48).192　　*l*(H6.8.48).192

Fig. IV.34. Operations acting on a square-tiled torus: vertex number differece: χ=0 (Euler characteristic of the torus).



H_{340}　　*dm*(H_{340}).506 {$4^3.4^5.4^6.4^7$}　　*m*(H_{340}).510

Fig. IV.35. Operations acting on a triple torus: vertex number differece: χ = -4; *g*=3.



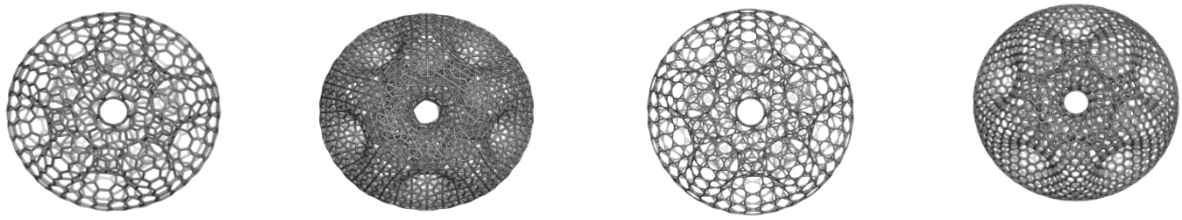
d(*s*(D)@*s*(D)7S.120).280　　$p_4(C_{280})$.820　　*l*(C_{280}).840

Fig. IV.36. Operations acting on a dodecahedral multi-torus: vertex number differece: χ = -20; *g*=11.

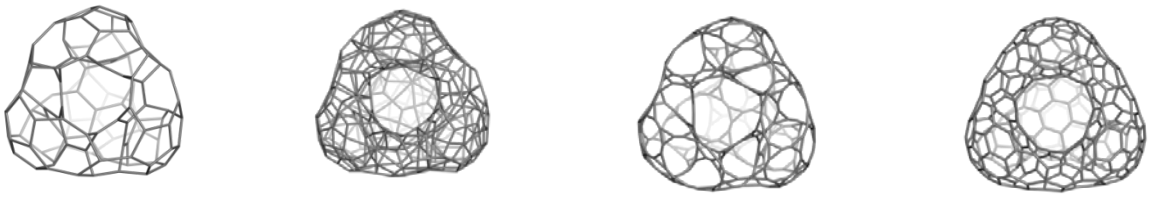
The figure count for the structures in Figures IV.34-36 is given in Table IV.29.

Table IV.29. Figure count for the objects in Figs. 2 to 4.

Structure	*v*	*e*	f_3	f_4	f_5	f_6	f_7	*f*	χ	*g*	*Diff*
H6.8	48	96	0	48	0	0	0	48	0	1	-
m(H6.8)	96	192	0	96	0	0	0	96	0	1	-
dm(H6.8)	96	192	0	96	0	0	0	96	0	1	0
H_{340}	340	510	0	0	12	118	36	166	-4	3	-
dm(H_{340})	506	1020	0	510	0	0	0	510	-4	3	-
m(H_{340})	510	1020	340	0	12	118	36	506	-4	3	- 4
C_{280}	280	420	0	0	0	0	120	120	-20	11	-
$p_4(C_{280})$	820	1680	0	840	0	0	0	840	-20	11	-
l(C_{280})	840	1260	0	0	0	280	120	400	-20	11	- 20



H.1220 *p4*H.1220.3640 *t*H.1220.3660 *l*H.1220.3660



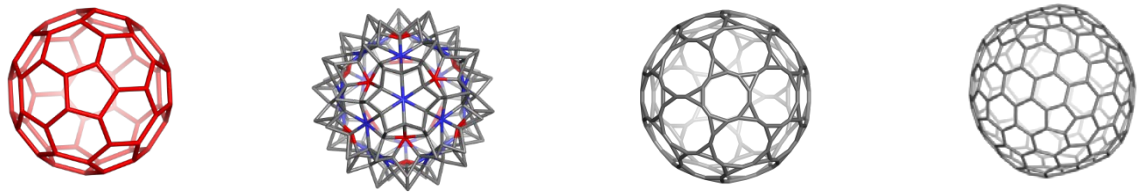
TTT57.104 p_4(TTT57.104).308 *t*(TTT57.104).312 *l*(TTT57.104).312

Fig. IV.37. Other example of pairs of operations

Table IV.30. Figure count in toroidal structures (Figures IV.37-38)

Structure	v	e	f3	f4	f5	f6	f7	f8	f10	f12	f14	f	χ	g
H.1220	1220	1830	0	0	0	470	120	0	0	0	0	590	-20	11
t H.1220	3660	5490	1220	0	0	0	0	0	0	470	120	1810	-20	11
p4 H.1220	3640	7320	0	3660	0	0	0	0	0	0	0	3660	-20	11

l H.1220	3660	5490	0	0	0	1690	120	0	0	0	0	1810	-20	11
C60	60	90	0	0	12	20	0	0	0	0	0	32	2	0
t C60	180	27	60	0	0	0	0	0	12	20	0	92	2	0
p4 C60	182	360	0	180	0	0	0	0	0	0	0	180	2	0
l C60	180	270	0	0	12	80	0	0	0	0	0	92	2	0
H68_48	48	96	0	48	0	0	0	0	0	0	0	0	0	1
t H68_48	192	288	0	48	0	0	0	0	0	0	0	0	0	1
p4 H68_48	192	384	0	192	0	0	0	0	0	0	0	0	0	1
l H68_48	192	288	0	48	0	0	0	48	0	0	0	96	0	1
TTT57	104	156	0	0	12	0	36	0	0	0	0	48	-4	3
t TTT57	312	468	104	0	0	12	0	0	0	6	42	152	-4	3
p4 TTT57	308	624	0	312	0	0	0	0	0	0	0	312	-4	3
l TTT57	312	468	0	0	12	104	36	0	0	0	0	152	-4	3



C_{60} $p_4(C_{60})$ $t(C_{60})$ $l(C_{60})$

Fig. IV.38. Operations acting on a C_{60}; vertex number differece: $\chi = 2$; g=0.

IV.7. Applications of Spectral Theory

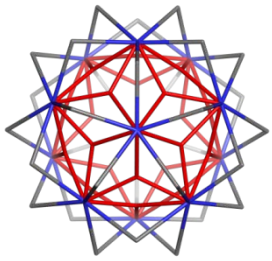
IV.7.1. Propellans

Rh_{30}.32 I@D@ID

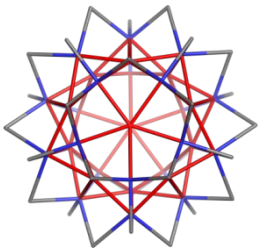
Fig. IV.39. Triacontahedron and three disconnected polyhedra composing propellanes.

All the rings in propellanes are rhombs. As a general property, all the vertex classes represent non-connected sets, thus the chromatic number equals the number of vertex classes. This property facilitates identification of vertex partitions as polyhedra and evaluation of their graph energy. A “binding” energy E_{bind} (in Beta units) can be calculated (see Table IV.31) for the parent graphs with respect to their independent partitions (*i.e.*, energy of composition, E_{compos}), by analogy with the quantum computations in molecular graphs.

$$E_{bind} = E - E_{compos} \qquad (10)$$

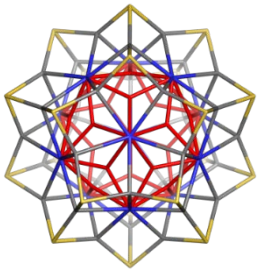


Rh_{30}_PrpA.62

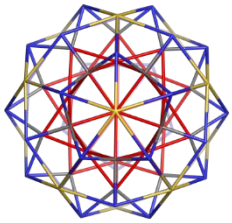


Rh_{30}_PrpB.62

Fig. IV.40. Propellanes of the 1st generation.

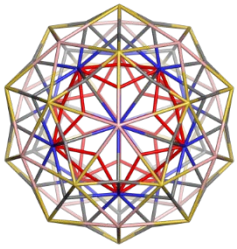


Rh_{30}_PrpAD.82

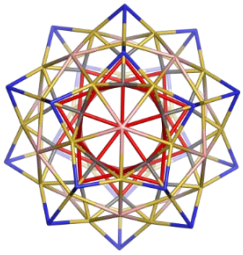


Rh_{30}_PrpBD.74

Fig.IV.41. Propellanes of the 2nd generation.



Rh_{30}_PrpACD.94



Rh_{30}_PrpBCD.94

Fig. IV.42. Propellanes of the 3rd generation.

Table IV.31. Propellane relatives of Rh_{30}.32; Graph energy; $E_{bind} = E - E_{compos}$.

	Cluster	Composition	*v*	E	E compos	E bind	λ max	λ min
1	Icosahedron		12	23.416			5	-2.236
2	Dodecahedron		20	29.416			3	-2.236
3	Icosidodecahedron		30	55.416			4	-2
4	Rh_{30}	I+D	32	47.896	52.832	-4.936	3.873	-3.873
5	Rh_{30}_Prp1A	I+D+ID	62	71.872	108.248	-36.376	5	-5
6	Rh_{30}_Prp1B	I+D+ID	62	87.314	108.248	-20.934	4.583	-4.583
7	Rh_{30}_Prp2A	I+2D+ID	82	129.43	137.664	-8.234	5.269	-5.269
8	Rh_{30}_Prp2B	2I+D+ID	74	132.828	131.664	1.164	5	-5
9	Rh_{30}_Prp3A	2I+2D+ID	94	154.906	161.08	-6.174	5.568	-5.568
10	Rh_{30}_Prp3B	2I+2D+ID	94	162.292	161.08	1.212	5.349	-5.349

IV.7.2. C_{40} Fullerenes

Graph energy in fullerene energy evaluation

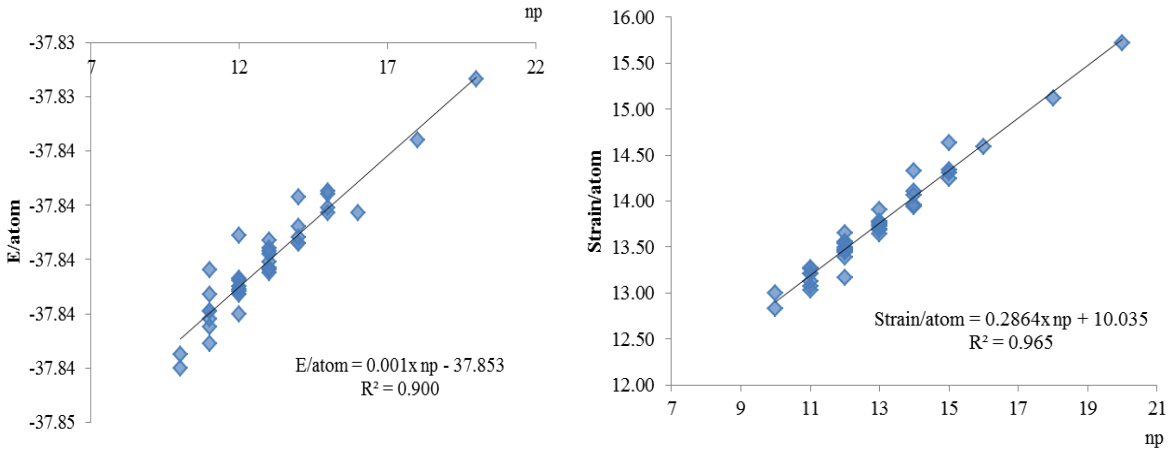


Fig.IV.43. Plots of total energy/atom (in au) and strain energy/atom (in kcal/mol) vs the number of fused pentagons n_p.

QSPR models (Table IV.35-36) can be derived in a variety of combinations; the goal was to show that the graph energy of these fullerenes can be used to evaluate the quantum calculated molecular energy. Table IV.34 lists several combinations, including the energy of graphs and remote graphs, with and without the n_p parameter.

Table IV.35. Best models in describing total energy E/atom.

X_1	X_2	X_3	R^2	s
n_p			0.900	0.000662
$SD_{E/atom}$			0.849	0.000813
LEig			0.780	0.000981
n_p	E_1		0.916	0.000616
	D_3		0.915	0.000617
n_p	E_1	LEig	0.924	0.000591
		D_3	0.921	0.000605
		C(Sh(D))	0.918	0.000614
$SD_{E/atom}$	LEig		0.897	0.000680
	D3D		0.882	0.000727
$SD_{E/atom}$	LEig	C(Sh(D))	0.918	0.000615
		D3D	0.907	0.000655

IV.8. Conclusions on Chapter IV

The following results were presented:

1. The smallest clusters of dimension/rank 4
2. Clusters of icosahedral symmetry
3. Clusters of octahedral symmetry
4. Cube and Rh_{12} networks
5. Clusters related to C_{60} fullerene
6. Energy of graphs in Propellanes
7. QSPR of C_{40} fullerenes using graph energies as descriptors

V. Personal Contributions (QSAR)

These QSAR studies were performed following Diudea's algorithm (Moldovan *et al*, 2008); it is based on the alignment of molecules over a hypermolecule (Balaban *et al*, 1980) and a correlation weighting procedure (Toropov *et al*, 2001,2002) coupled with a predictive validation of the model descriptors within similarity clusters (Willett, 1998) performed for each molecule in the test set.

V.1. QSAR study on phenothiazines

Phenothiazine is an organic heterocyclic compound, of the class of thiazines, with the brute formula $S(C_6H_4)_2NH$. Derivatives of phenothiazine revolutionized the field of psychiatry and allergy treatment.

Table V.1. Studied phenothiazines with their name, CID, properties logP and LD_{50}.

Mol	CID	Name	Canonical Smiles	logP	LD50 mg/kg
1	2726	chlorpromazine	CN(C)CCCN1C2=CC=CC=C2SC3=C1C=C(C=C3)Cl	5.41	14
2	2801	clomipramine	CN(C)CCCN1C2=CC=CC=C2CCC3=C1C=C(C=C3)Cl	5.19	150
3	2995	desipramine	CNCCCN1C2=CC=CC=C2CCC3=CC=CC=C31	4.90	85
4	3089	fonazine/dimetothiazine	CC(CN1C2=CC=CC=C2SC3=C1C=C(C=C3)S(=O)(=O)N(C)C)N(C)C	3.34	190
5	3781	isothipendyl	CC(CN1C2=CC=CC=C2SC3=C1N=CC=C3)N(C)C	3.66	62
6	4066	mequitazine	C1CN2CCC1C(C2)CN3C4=CC=CC=C4SC5=CC=CC=C53	4.70	54
7	4744	perazine	CN1CCN(CC1)CCCN2C3=CC=CC=C3SC4=CC=CC=C42	4.10	185
8	4747	periciazine	C1CN(CCC1O)CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)C#N	3.52	115
9	4748	perphenazine	C1CN(CCN1CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)Cl)CCO	4.20	64
10	4917	prochlorperazine	CN1CCN(CC1)CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)Cl	4.88	120
11	4926	promazine	CN(C)CCCN1C2=CC=CC=C2SC3=CC=CC=C31	4.55	140
12	4927	promethazine	CC(CN1C2=CC=CC=C2SC3=CC=CC=C31)N(C)C	4.81	124
13	5452	thioridazine	CN1CCCCC1CCN2C3=CC=CC=C3SC4=C2C=C(C=C4)SC	5.90	65
14	5566	trifluoperazine	CN1CCN(CC1)CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)C(F)(F)F	5.03	120
15	6075	mepazine/pecazine	CN1CCCC(C1)CN2C3=CC=CC=C3SC4=CC=CC=C42	5.60	140
16	6077	acetylpromazine	CC(=O)C1=CC2=C(C=C1)SC3=CC=CC=C3N2CCCN(C)C	4.20	350
17	6761	pipamazine	C1CN(CCC1C(=O)N)CCCN2C3=CC=CC=C3SC4=C2C=C(C=C4)Cl	4.40	80
18	10646	pyrathiazine	C1CCN(C1)CCN2C3=CC=CC=C3SC4=CC=CC=C42	4.70	190
19	14670	prothypendyl	CN(C)CCCN1C2=CC=CC=C2SC3=C1N=CC=C3	3.40	135
20	14677	methdilazine	CN1CCC(C1)CN2C3=CC=CC=C3SC4=CC=CC=C42	5.23	183
21	16414	7-hidroxyclorpromazine	CN(C)CCCN1C2=C(C=C(C=C2)O)SC3=C1C=C(C=C3)Cl	4.80	119
22	19396	oxomemazine	CC(CN1C2=CC=CC=C2S(=O)(=O)C3=CC=CC=C31)CN(C)C	3.40	185
23	19675	piperacetazine	CC(=O)C1=CC2=C(C=C1)SC3=CC=CC=C3N2CCCN4CCC(CC4)CCO	4.21	98
24	65535	diethazine	CCN(CC)CCN1C2=CC=CC=C2SC3=CC=CC=C31	4.90	225
25	65750	chlorproethazine	CCN(CC)CCCN1C2=CC=CC=C2SC3=C1C=C(C=C3)Cl	5.90	90
26	68223	fenethazine	CN(C)CCN1C2=CC=CC=C2SC3=CC=CC=C31	4.20	115
27	69500	difazin	CCN(CC)CC(=O)N1C2=CC=CC=C2SC3=CC=CC=C31	3.80	210
28	70413	opromazine	CN(C)CCCN1C2=CC=CC=C2S(=O)C3=C1C=C(C=C3)Cl	3.90	163
29	72287	levomepromazine	C[C@@H](CN1C2=CC=CC=C2SC3=C1C=C(C=C3)OC)CN(C)C	4.68	58.5
30	94280	dimetacrine	CC1(C2=CC=CC=C2N(C3=CC=CC=C31)CCCN(C)C)C	4.96	206

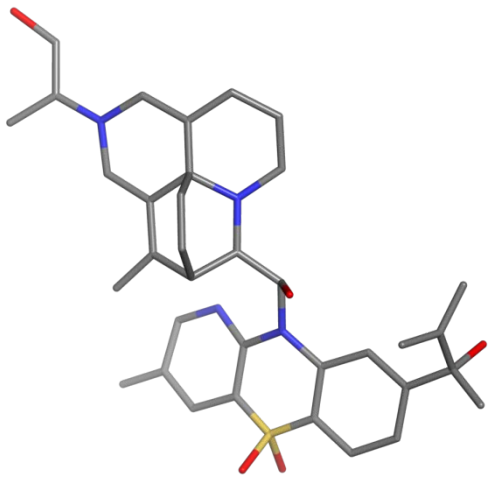


Fig. V.1. Hypermolecule comprising the features of the dataset

Basic equations that describe the relationships between values of property or biological activity of compounds and their structures were obtained:

$logP = 59.096 + SD_{logP}$; n=30; R^2=0.946; s=0.165; F=488.078 (1)

$LD_{50} = 9113.289 + SD_{LD50}$; n=28; R^2=0.956; s=13.964; F=566.487 (2)

(molecules 1 and 23 were outliers, residual value>2s)

QSAR models (for case log P)

The models were performed on the training set (structures 11-30) and the best results (in decreasing order of R^2) are listed below in Tables V.3 and V.4 (Pîrvan-Moldovan, 2016).

Monovariate regression $logP = 57.266 + 0.966 \times SD_{logP}$ (3)
n=20; R^2=0.946; s=0.172; F=317.170

Bivariate regression $logP = 57.299 + 0.968 \times SD_{logP} + 3.54 \times 10^{-5} \times IP[CfMax]$ (4)
n=20; R^2=0.949; s=0.173; F=157.155

Trivariate regression $logP = 56.341 + 0.948 \times SD_{logP} + 1.26 \times 10^{-} \times IP[CfMin] + 0.097 \times Chem.pot.$ (5)
n=20; R^2=0.951; s=0.174; F=103.820

QSAR models (for case LD_{50}) (https://www.drugbank.ca/)

The models were performed on the training set and the best results (in decreasing order of R^2) are listed below in Tables V.5 and V.6.

Monovariate regression $LD_{50}=9444.65+1.037\times SD_{LD50}$ (6)
n=19; R^2=0.940; s=13.534; F=265.126)

Bivariate regression $LD_{50}=9415.945+1.033\times SD_{LD50}-0.0266\times IE[CfMax]$ (7)
n=19; R^2=0.943; s=13.607; F=131.573)

Trivariate regression $LD_{50}=9606.267+1.055\times SD_{LD50}-0.2024\times IE[CfMax]+0.047\times Distance$ (8)
n=19; R^2=0.947; s=13.498; F=89.557

Models validation

(a) Leave-one-out

The performances in leave-one-out analysis (Jäntschi, 2005) related to the models listed as best in Tables are shown in Tables V.7 and V.8.

Table V.7. Leave-one-out analysis for the best logP models

	Descriptors	Q^2	R^2- Q^2
1	SD_{logP}	0.938	0.008
2	SD_{logP}, IP[CfMax]	0.928	0.021
3	SD_{logP}, IP[CfMin], Chem.pot.	0.910	0.041

Table V.8. Leave-one-out analysis for the best LD_{50} models

	Descriptors	Q^2	R^2- Q^2
1	SD_{LD50}	0.931	0.009
2	SD_{LD50}, IE[CfMax]	0.925	0.018
3	SD_{LD50}, IE[CfMax], Distance	0.919	0.030

(b) External Validation

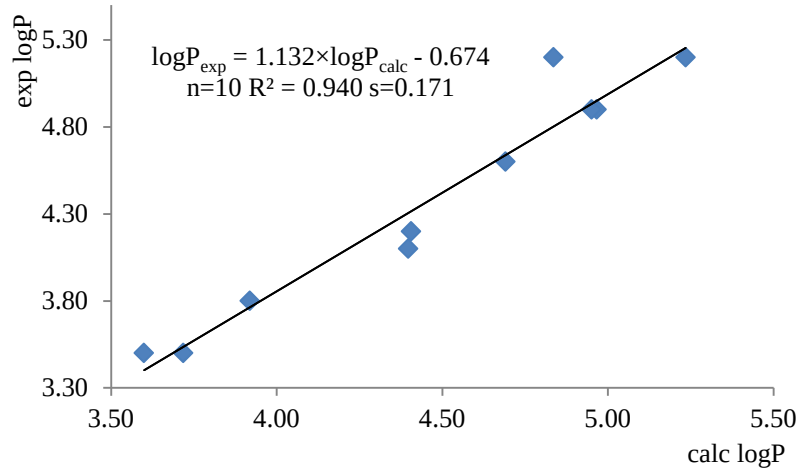


Fig.V.3. The plot $logP_{exp}$ vs. $logP_{calc.}$ for the test set (external validation).

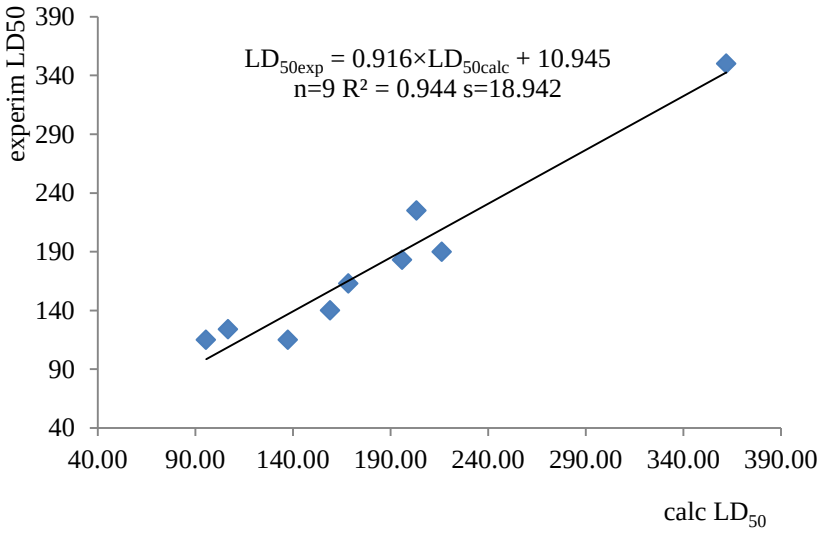


Fig.V.4. The plot LD_{50exp} vs. $LD_{50calc.}$ for the test set (external validation)

From Figures V.3 and V.4 one can see that our models show a good predictive ability.

(c) Similarity Cluster Validation

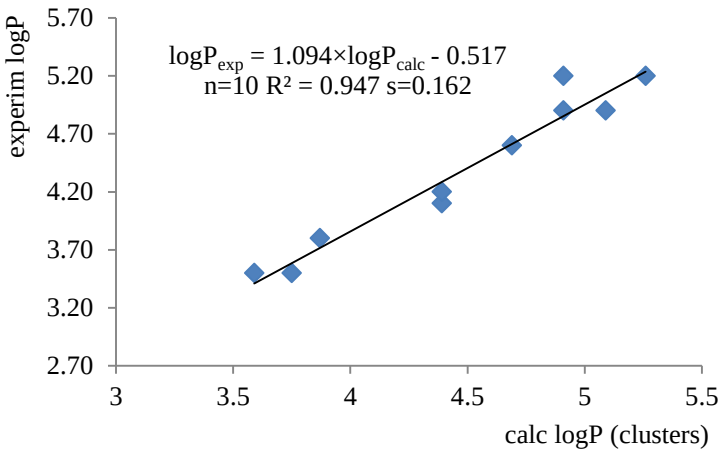


Fig.V.5. The plot logP$_{exp}$ vs. logP$_{calc}$ (by clusters of similarity) for the test set

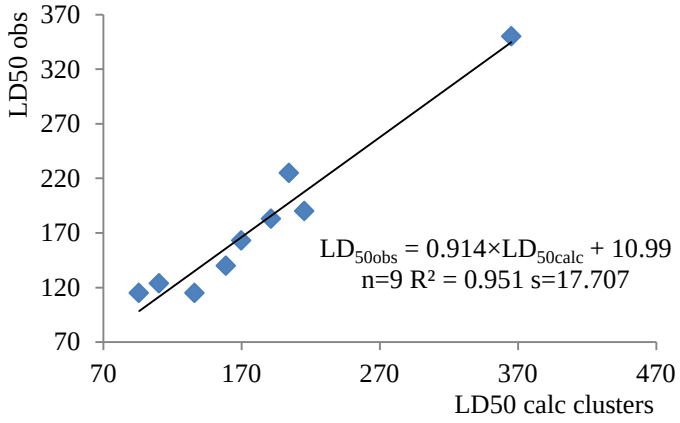


Fig. V.6. The plot LD$_{50}$ vs. LD$_{50calc}$ (by clusters of similarity) for the test set

V.2. QSAR study on triptan class

Sumatriptan, $C_{14}H_{21}N_3O_2S$, is a synthetic drug included in the triptan class used for the treatment of migraine.

Table V.9. List of studied sumatriptans with their CID and property logP (https://pubchem.ncbi.nlm.nih.gov)

Molecule	XLogP3	CID	Canonical SMILES
1	0.9	5358	CNS(=O)(=O)CC1=CC2=C(C=C1)NC=C2CCN(C)C
2	2	4440	CNS(=O)(=O)CCC1=CC2=C(C=C1)NC=C2C3CCN(CC3)C
3	4.1	77993	CN1CCC[C@@H]1CC2=CNC3=C2C=C(C=C3)CCS(=O)(=O)C4=CC=CC=C4
4	1.2	10043491	CN1CCCN(S1(=O)=O)CC2=CC3=C(C=C2)NC=C3CCN(C)C
5	1.9	10021702	CCN1CCN(S1(=O)=O)CCC2=CC3=C(C=C2)NC=C3CCN(C)C
6	2.2	34855	CC1=CC2=C(C=C1)NC=C2CN(C)C
7	1.2	10257	CN(C)CCC1=CNC2=C1C=C(C=C2)O
8	2.7	10340828	CN(C)CCC1=CNC2=C1C=C(C=C2)N3CCN(S3(=O)=O)CC4=CC=CC=C4
9	-1.2	439280	C1=CC2=C(C=C1O)C(=CN2)C[C@@H](C(=O)O)N
10	0.1	11722814	CN1CCN(S1(=O)=O)CC2=CC3=C(C=C2)NC=C3CCN
11	1.4	10404770	CN(C)CCC1=CNC2=C1C=C(C=C2)CCN3CCNS3(=O)=O
12	1.9	9998879	CC(C)N1CCN(S1(=O)=O)CC2=CC3=C(C=C2)NC=C3CCN(C)C
13	1.6	123606	CN(C)CCC1=CNC2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3
14	2.3	9802530	CN(C)CCC1=CNC2=C1C=C(C=C2)OS(=O)(=O)C(F)(F)
15	3	10618751	CCN(CC)CCC1=CNC2=C1C=C(C=C2)OS(=O)(=O)C(F)(F)F
16	0.5	4713248	CNS(=O)(=O)C1=CC2=C(C=C1)NC=C2CCN(C)C
17	3.9	9954663	C1CCN(C1)CCC2=CNC3=C2C=C(C=C3)C4=CCN(CC4)S(=O)(=O)C5=CC=CC=C5
18	1.5	18423663	CNS(=O)(=O)CC1=CC2=C(C=C1)NC=C2CC3CCCN3C
19	1	18423665	CNS(=O)(=O)CC1=CC2=C(C=C1)NC=C2CC3CCCN3
20	4.6	11177383	CC(C)C1=CC=C(C=C1)S(=O)(=O)NC2=CC3=C(C=C2)NC=C3CC4CCCN4C
21	1.5	12822482	C1=CC=C(C=C1)S(=O)(=O)NCC2=CC3=C(C=C2)NC=C3CCN
22	2.3	53644294	C1CN(CC1CNCC2=CC=CC=C2)CCC3=CNC4=C3C=C(C=C4)CS(=O)(=O)N
23	4.3	19422723	CN1CCCC1CC2=CNC3=C2C=C(C=C3)C(=C)S(=O)(=O)C4=CC=CC=C4
24	2.9	67923722	CNS(=O)(=O)C(CC1=CC=CC=C1)C2=CC3=C(C=C2)NC=C3CCN(C)C
25	3	44400804	C1CC(NC1)CC2=CNC3=C2C=C(C=C3)NS(=O)(=O)C4=CC=CC=C4
26	2.2	13475733	C1=CC=C(C=C1)CNS(=O)(=O)CCC2=CC3=C(C=C2)NC=C3CCN
27	1.4	13286052	C1=CC=C(C=C1)CNS(=O)(=O)CC2=CC3=C(C=C2)NC=C3CCN
28	1.5	13286055	C1=CC=C(C=C1)NS(=O)(=O)CC2=CC3=C(C=C2)NC=C3CCN
29	1.4	12822479	C1=CC=C(C=C1)CS(=O)(=O)NCC2=CC3=C(C=C2)NC=C3CCN
30	3.3	12082798	CN1CCC(=CC1)C2=CNC3=C2C=C(C=C3)CCS(=O)(=O)N(C)CC4=CC=CC=C4
31	2.4	10571277	CNS(=O)(=O)CC1=CC2=C(C=C1)NC=C2CCN(C)CC3=CC=CC=C3
32	5	10203969	C1CCN(C1)CCC2=CNC3=C2C=C(C=C3)NS(=O)(=O)C4=CC=C(C=C4)C5=CC=CC=
33	2.1	102242445	CCC1=CNC2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3
34	3.6	89221361	CN1CCCC1C2=CNC3=C2C=C(C=C3)/C=C/S(=O)(=O)C4=CC=CC=C4
35	1.8	77350860	CNS(=O)(=O)CC1=CC2=C(C=C1)NC=C2CCC3CCCN3C
36	2.8	76115652	CCCC1CCN(C1)CC2=CNC3=C2C=C(C=C3)CS(=O)(=O)N(C)C
37	3	71510938	CNCCC1=CN(C2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3)CC4=CC=CC=C4
38	1.5	11809037	CN1CCC(CC1)C2=CNC3=C2N=C(C=C3)OS(=O)(=O)C
39	1.1	71315712	CNCCC1=CNC2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3
40	3.6	70946434	CN1CCCC1C2=CNC3=C2C=C(C=C3)CCS(=O)(=O)C4=CC=CC=C4
41	2	70836771	CN(CCC1=CNC2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3)CF
42	1.9	70932894	CC(C1=CNC2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3)N(C)C
43	1.7	67834586	CC(C1=CC2=C(C=C1)NC=C2CC3CCCN3)S(=O)(=O)NC
44	1.6	67767267	CC1(CN(S(=O)(=O)N1)CC2=CC3=C(C=C2)NC=C3CCN(C)C)C
45	4.8	67424557	C1CC(N(C1)CI)CC2=CNC3=C2C=C(C=C3)CCS(=O)(=O)C4=CC=CC=C4
46	5.4	67125361	CCCC(CC1=CC2=C(C=C1)NC=C2CC3CCCN3C)S(=O)(=O)C4=CC=CC=C4
47	1.2	66617669	CC(C)(CC1=CNC2=C1C=C(C=C2)CS(=O)(=O)N3CCCC3)N
48	0.6	54293224	C1CC(NC1)CC2=CNC3=C2C=C(C=C3)CS(=O)(=O)N
49	0.8	54277349	CNS(=O)(=O)CC1=CC2=C(C=C1)NC=C2C3=CCN(CC3)C
50	1.8	54043760	CNS(=O)(=O)CCC1=CC2=C(C=C1)NC=C2C[C@H]3CCCN3
51	3.8	44400787	CC1=CC=C(C=C1)S(=O)(=O)NC2=CC3=C(C=C2)NC=C3CC4CCCN4C
52	0.6	10448821	CC1(CN(S(=O)(=O)N1)CC2=CC3=C(C=C2)NC=C3CCN)C
53	2.4	23192135	CN1CCCC1CC2=CNC3=C2C=C(C=C3)CCS(=O)(=O)N(C)C
54	1.2	22924406	CN1CCC(=CC1)C2=CNC3=C2C=C(C=C3)CCS(=O)(=O)N
55	1.8	22371684	CC(C1=CC2=C(C=C1)NC=C2CC3CCCN3C)S(=O)(=O)N
56	2.8	21852525	CC(C)(C)NS(=O)(=O)CC1=CC2=C(C=C1)NC=C2CC3CCCN3C
57	1.6	20066734	CNS(=O)(=O)CCC1=CC2=C(C=C1)NC=C2C3=CCN(CC3)C
58	1.9	19970226	CN1CCCC1CC2=CNC3=C2C=C(C=C3)CCS(=O)(=O)N
59	4	19970209	CC(C1=CC2=C(C=C1)NC=C2CC3CCCN3C)S(=O)(=O)C4=CC=CC=C4
60	3.6	19041599	C1CC(NC1)CC2=CNC3=C2C=C(C=C3)CCS(=O)(=O)C4=CC=CC=C4

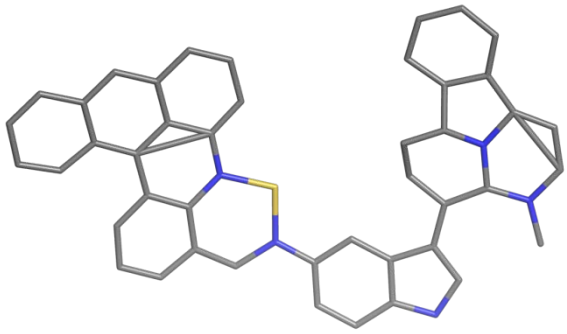


Fig.V.7.a. HM designed for molecules that have S at 2 atoms distance from indole

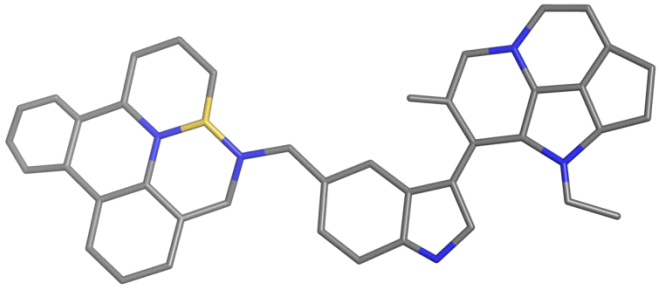


Fig. V.7.c. HM designed for molecules that have S at 3 atoms distance from indole

Basic equations that describe the relationships between values of property of 30, respectively 20 compounds and their structures were obtained; molecules 5, 9 and 10 are outliers.

$$logP=-15.734+0.999\times SD_{logP};\qquad n=30,\ R^2=0.927,\ s=0.321,\ F=355.73 \qquad (13)$$

$$logP=13.628+SD_{logP};\qquad n=20,\ R^2=0.9998,\ s=0.017,\ F=119140 \qquad (14)$$

QSAR models in set "2" (for case logP)

The models were performed on the training set and the best results (in decreasing order of R^2) are listed in Table V.12.

QSAR models:

Monovariate regression	$logP= -15.294+0.978\times SD_{logP}$	(15)
	$n=21, R^2=0.952, s=0.280$	
Bivariate regression	$logP= -16.582+1.124\times SD_{logP} -0.029\times N$	(16)
	$n=21, R^2=0.966, s=0.242$	
Trivariate regression	$logP= -16.582+1.154\times SD_{logP} -0.032\times N-0.431\times HOMO$	(17)
	$n=21, R^2=0.969, s=0.237$	

Models validation

a. Internal validation

The performances in leave-one-out analysis related to the models listed as best are shown in Table V.13.

Table V.13. Leave-one-out analysis for the best logP models

	Descriptors	R^2	Q^2	R^2- Q^2
1	SD_{logP}	0.952	0.944	0.009
2	SD_{logP}, N	0.966	0.955	0.012
3	SD_{logP}, N, HOMO	0.969	0.956	0.014

b. External Validation

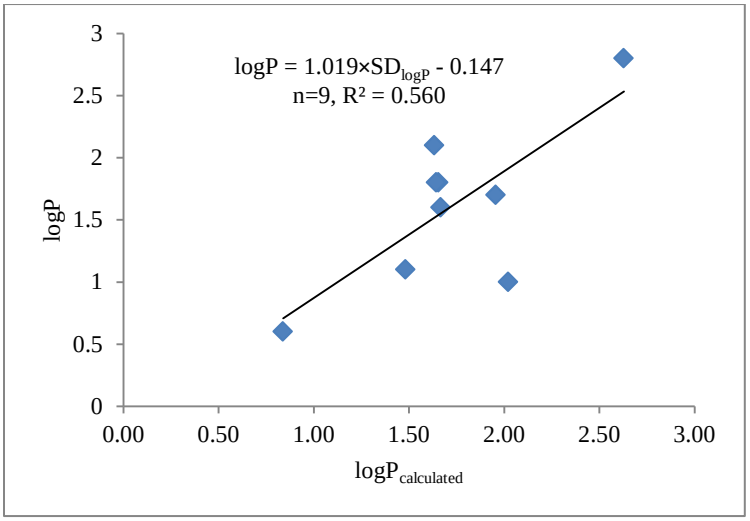


Fig.V.8. The plot $logP_{exp}$ vs. $logP_{calc.}$ f or the test set (external validation).

c. Similarity Cluster Validation

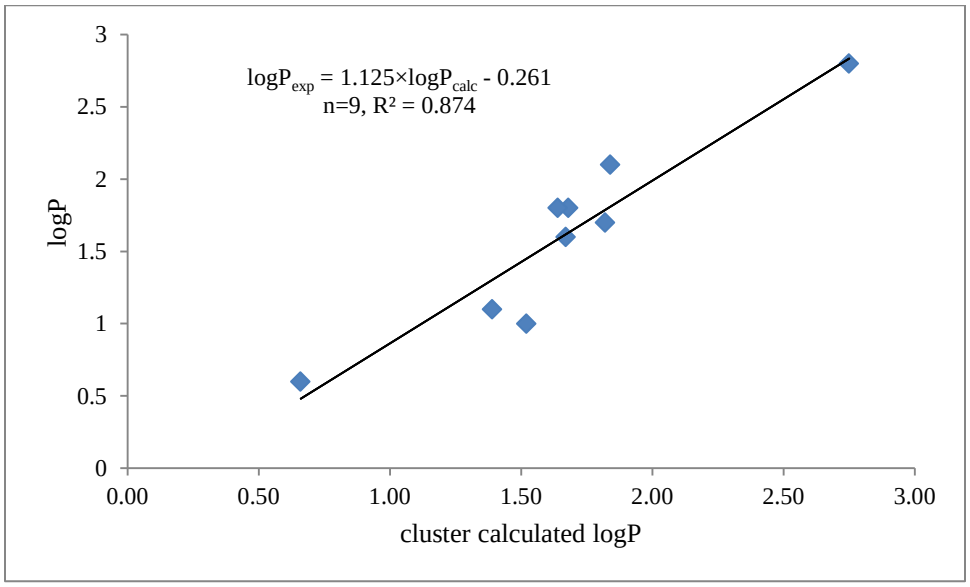


Fig. V.9. The plot $logP_{exp}$ vs. $logP_{calc.}$ for the test set (**similarity** cluster validation).

QSAR models in set "2" for case Binding Affinity

Ligand-protein binding is based on intermolecular forces, like ionic bonds, hydrogen bonds and Van der Waals forces. After association of docking could take place the dissociation, being a reversible process. The rate of binding is called affinity. BA values were obtained by molecular docking procedure using AutoDock 4.2.6 (Morris *et al*, 2009

Basic equation that describes the relationships between 32 compounds Binding Affinity values and their structures is obtained.

$$\text{Binding Affinity} = -11.827 + SD_{BA} \quad \text{(kcal/mol)}; \qquad n=32,\ R^2=0.915,\ s=0.253,\ F=324.639 \tag{18}$$

Model validation

After similarity analysis was determined molecules cluster for the chosen ligand to test. Each cluster includes the same number of molecules. In Table V.17 are some similarity results obtained using TOPOCLUJ software.

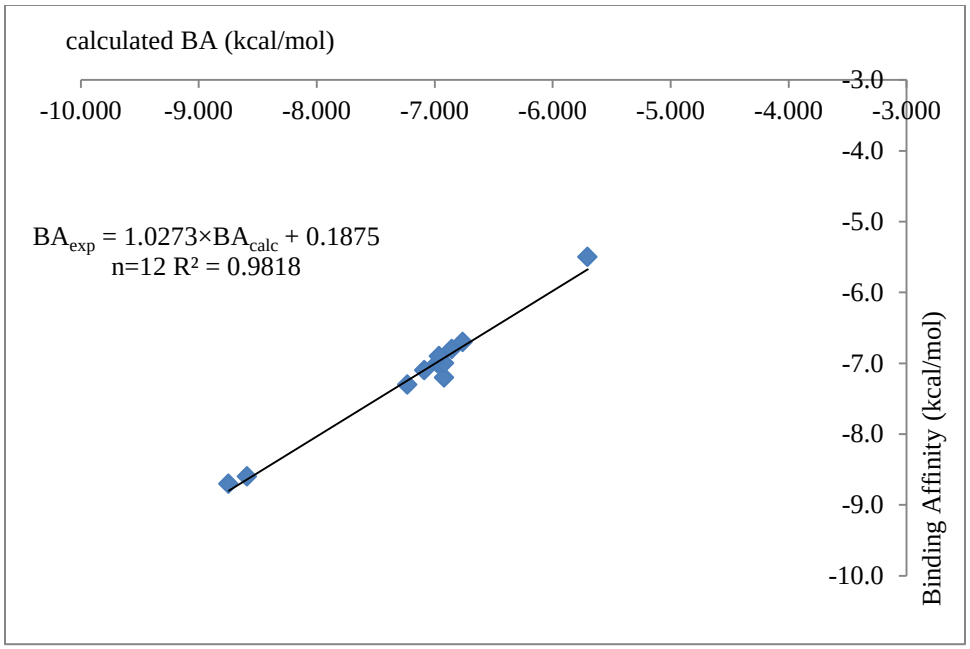


Fig.V.10. The plot BA_{exp} vs. $BA_{calc.}$ for the test set (**similarity** cluster validation).

The plot plot BA_{exp} vs. $BA_{calc.}$ for the test set (Fig.V.10) shows a good correlation between experimental data and calculated ones.

Predicting Binding Affinity for new ligands

Three new molecules for each set have been proposed, each of them was included in its own set for calculating SD_{logP} (Table V.25). Every new ligand was compared with its set, different clusters were made, different predicting models were established.

Table V.25. New proposed ligands

New ligand	Average BA (kcal/mol)	SD_{BA}	Predicted BA (kcal/mol)	Docking BA (kcal/mol)	Average BA of conformes (kcal/mol)
2a	-7.30	9.494	-8.26	-8.7	-7.9
2b	-7.23	9.297	-8.27	-8.8	-8.3
2c	-7.18	9.829	-7.82	-8.9	-8.5
3a	-7.46	6.998	-7.49	-8.4	-7.8
3b	-7.49	7.004	-7.52	-8.0	-7.6
3c	-7.46	6.480	-8.02	-8.1	-7.8

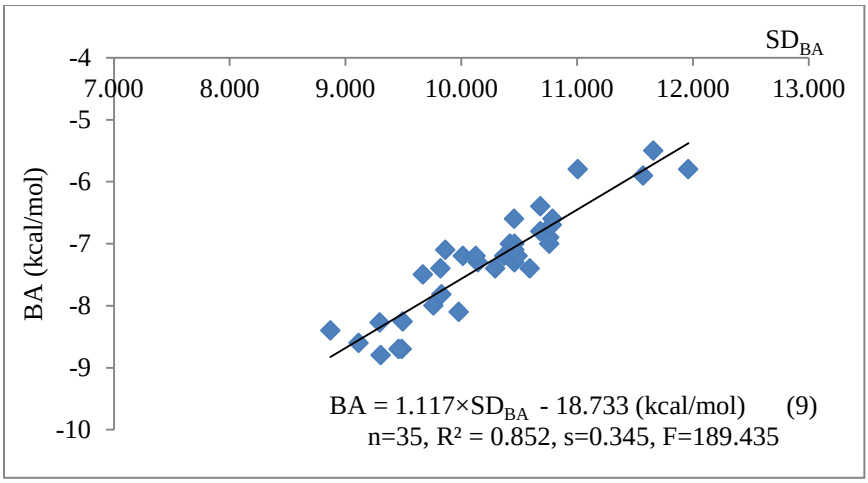


Fig.V.13. The plot $BA_{docking}$ vs. SD_{BA} for the set "2"

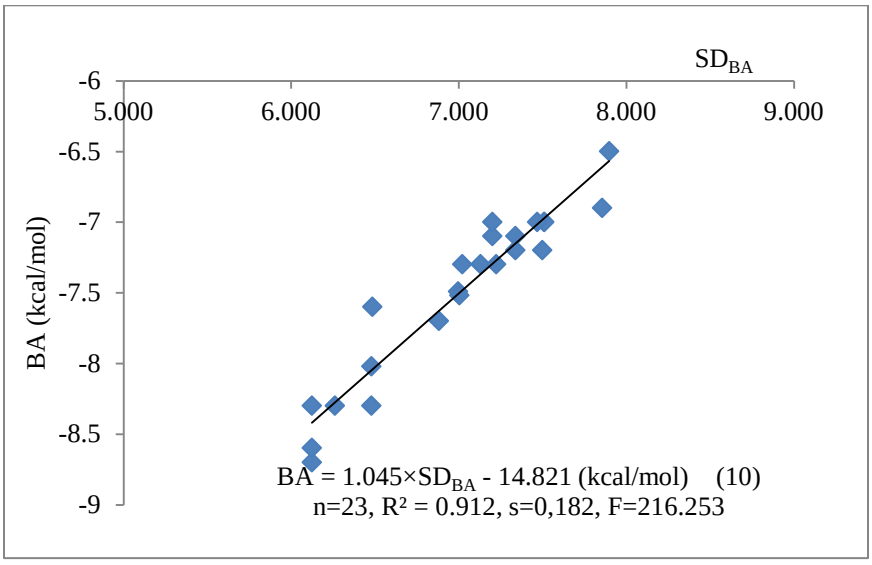


Fig.V.14. The plot $BA_{docking}$ vs. SD_{BA} for the set "3"

Concluding, it is possible to predict activities of untested congeneric molecules if the study is based on good experimental results.

V.3. QSAR study on sildenafil derivatives

Sildenafil, $C_{22}H_{30}N_6O_4S$, is a strong and selective inhibitor of cGMP-specific phosphodiesterase type 5 (PDE5), an enzyme that sustains degradation of cyclic guanosine monophosphate (cGMP), which adjusts blood flow in the corpus cavernosum.

A 40 molecules set was downloaded from a (https://pubchem.ncbi.nlm.nih.gov) database, Table V.26.

Table V.26. List of studied sildenafil derivatives with their CID and properties logP and TPSA.

#	CID	Canonical SMILES	XLogP	TPSA (A^2)
1	5212	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC)C	1.5	118
2	44440137	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCC(CC4)P(=O)(OCC)OCC)OCCC)C	2.7	150
3	44440128	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)CP(=O)(O)OCC)OCCC)C	-1.1	164
4	44391946	CCC1=C2C(=NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC)N(N1)C5CCCC5	2.6	115
5	110634	CCCC1=NC(=C2N1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)CC)OCC)C	2.5	118
6	44391895	CCCN1C2=NC(=NC(=O)C2=C(N1)CC)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC	2.2	115
7	44402347	CCOC1=C(C=C(C=C1)S(=O)(=O)N2CCN(CC2)C)C3=NC(=O)C4=C(N3)N5C=CC=C(C5=N4)C	2.5	117
8	44402480	CCOC1=C(C=C(C=C1)S(=O)(=O)N2CCN(CC2)C)C3=NC(=O)C4=C(N3)N5C=C(C=CC5=N4)C	2.5	115
9	44402581	CCN1CCN(CC1)S(=O)(=O)C2=CC(=C(C=C2)OCC)C3=NC(=O)C4=C(N3)N5C(=CC=CC5=N4)C	2.9	117
10	45266248	CCOC1=C(C=C(C=C1)S(=O)(=O)N2CCN(CC2)C)C3=NC(=O)C4=C(N3)N5C=CC=CC5=N4	2.2	117
11	9913987	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCC(CC4)C(=O)N)OCC)C	1	157
12	44381943	CCCN1C=NC2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCCC	1.8	118
13	45267330	CCCC1=C(C=C(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC)O)OC	3.2	129
14	72543811	CCCC1=NN(C2=C1N=C(N=N2)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC)C	1.4	115
15	118728651	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)NCCCCO)OCC)C	1.5	143
16	118728655	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)NCCN(CC)CC)OCC)C	2.3	126
17	118728656	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)NCCCN(C)C)OCC)C	1.9	126
18	9935230	CCCC1=NC(=C2N1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC)CC	2.5	118
19	10227317	CCCC1=C2NC(=NC(=O)N2C(=N1)C)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC	2.1	118
20	10277925	CCN1CCN(CC1)S(=O)(=O)C2=CN=C(C(=C2)C3=NC(=O)C4=NN5CCCCC5=C4N3)OCC	0.6	130
21	11540832	CCC1=C2C(=O)N=C(NN2C(=N1)C3CCCC3)C4=C(C=CC(=C4)S(=O)(=O)N5CCN(CC5)C)OCC	3	118
22	9955904	CCCN1C=NC2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC	1.2	118
23	9956558	CCCN1C2=NC(=NC(=O)C2=C(N1)C)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC	1.8	115
24	24859502	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C5CCCCC5)OCC)C	3.3	118
25	70687499	CCOC1=C(C=C(C=C1)S(=O)(=O)N2CCN(CC2)C)C3=NC(=O)C4=C(N3)C=NN4C	0.3	118
26	9845589	CCCC1=NN(C2=C1NC(=NC2=O)C3=CC(=CC=C3)S(=O)(=O)N4CCN(CC4)C)C	1.1	108
27	1896867	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCCCC4)OCC)C	2.5	114
28	53956764	CCOC1=C(C=C(C=C1)S(=O)(=O)N2CCN(CC2)C)C3=NC(=O)C4=C(N3)C(=NN4C)	0.7	118
29	10072962	CCCC1=C2C(=NN1)C(=O)N=C(N2)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC	1.5	128
30	10141109	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CC(NC(C4)C)C)OCC)C	1.9	126
31	10228242	CCCC1=NN(C2=C1NC(=NC2=S)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C)OCC)C	2.1	133
32	12018718	CCCC1=NC(=C2N1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCNCC4)OCC)C	1.6	126
33	24756844	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)CC)OCC)C	1.8	118
34	56841591	CCCC1=NN(C2=C1NC(=NC2=S)C3=C(C=CC(=C3)S(=O)(=O)N4CC(NC(C4)C)C)OCC)C	2.5	141
35	25209618	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCN(CC4)C5CCCC5)OCC)C	2.8	118
36	102582026	CCCC1=NN(C2=C1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N4CCNCC4)OCCC)CC	1.8	126
37	23449792	CCCCOC1=NC=C(C=C1C2=NC(=O)C3=NNC(=C3N2)CC)S(=O)(=O)N4CCN(CC4)CC	1.6	141
38	22893606	CCCCCC1=NC(=C2N1NC(=NC2=O)C3=C(C=CC(=C3)S(=O)(=O)N(CC)CC)OCC)C	4.1	115
39	23522695	CCCC1=C2NC(=NC(=O)N2C(=N1)C)C3=C(C=CC(=C3)S(=O)(=O)N4CCCCC4)OCC	3.1	114
40	23522716	CCCC1=C2NC(=NC(=O)N2C(=N1)C)C3=C(C=CC(=C3)S(=O)(=O)NO)OCC	1.6	143

Diudea's algorithm is followed in this study, too.

- Downloading from a database of molecular structures; spliting in training set and test set.
- Geometry optimization (Energy and partial charges calculation)
- Global/local descriptors/indices calculation
- Hypermolecule and its vectors calculation (more than one alignments)
- Correlation weighting and data reduction
- Sum descriptor SD calculation
- QSAR Models (Toplis & Costello; Chance correlation)
- Internal&External validation
- Similarity clusters calculation
- Prediction/validation by similarity clusters

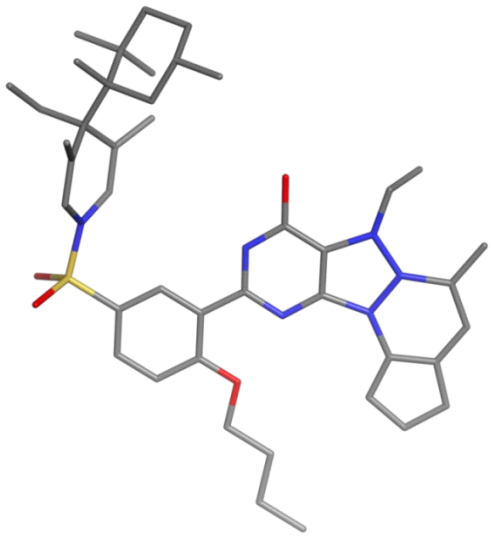


Fig.V.14.b. Hypermolecule comprising the features of the dataset

Basic equations on entire molecules set was obtained. From logP point of view, molecule 32 is an outlier.

$$logP=16.901+SD_{logP}; \quad n=39,\ R^2=0.934,\ s=0.245,\ F=525.792 \quad (21)$$

$$TPSA=207.877+0.989\times SD_{TPSA}; \quad n=37,\ R^2=0.925,\ s=0.323,\ F=433.889 \quad (22)$$

QSAR models with 1 variable

In training set basic equations were obtained.

$logP=17.121+1.0108\times SD_{logP}$; n=26, R^2=0.953, s=0.228, F=485.105 (23)

$TPSA=210.063+0.987\times SD_{TPSA}$; n=26, R^2=0.953, s=2.954, F=486.036 (24)

Conclusions to Tables V.29-30: Although R^2 increases, standard error is not decreasing. It is obvious that standard error (monovariate model) < standard error (bivariate model)< standard error (trivariate model), so the most simple models are to be used in future predictions (Eq. 23 and 24).

Table V.31. Statistical parameters of multilinear regression in logP study

logP Descriptors			R^2	s	F
SD_{logP}			0.9529	0.2279	485.1048
SD_{logP}	LUMO		0.9535	0.2311	235.9656
SD_{logP}	LUMO	Detour	0.9537	0.2359	151.0588

Table V.32. Statistical parameters of multilinear regression in TPSA study

TPSA Descriptors			R^2	s	F
SD_{TPSA}			0.9529	2.9541	486.0361
SD_{TPSA}	HOMO		0.9546	2.9634	241.9298
SD_{TPSA}	HOMO	logWkOp[Adj.Det.D3D]	0.9554	3.0033	157.1547

Models validation

Internal validation **LOO**

Q^2 is the corresponding R^2 to the prediction**.**

Table V.33. Leave-one-out results

Descriptor	R^2	Q^2	R^2-Q^2
SD_{logP}	0.9529	0.9468	0.0061
SD_{TPSA}	0.9529	0.9481	0.0048

External validation

It makes up clusters of similarity (within the molecules of the learning set) for each of the test-molecules; it generates models for the test clusters (excluding the leaders of similarity); it predicts the activity values for the test-molecules (individual, on each similarity cluster) (Table V.34).

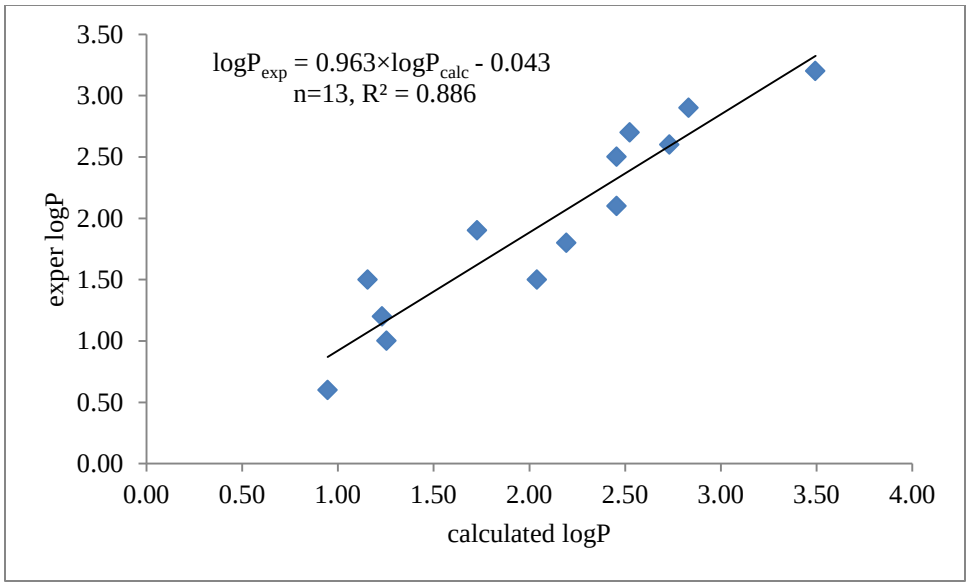


Fig. V.15. The plot logP$_{exp}$ vs. logP$_{calc.}$ for the test set (external validation).

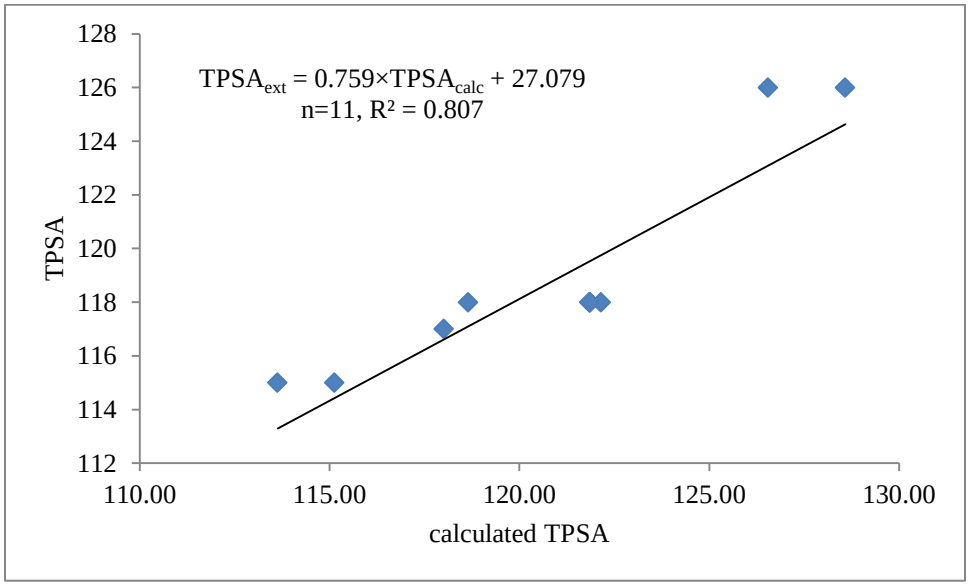


Fig. V.16. The plot TPSA$_{ext}$ vs. TPSA$_{calc.}$ for the test set (external validation).

Similarity Cluster Validation

The principle of clusters composition was the same. Only molecules with similarity over 80% were retained. Based on similarity clusters, calculated values of logP and TPSA were obtained (Table V.36-37).

Figures V.17-18 show the degree of correlation between experimental values of studied properties and calculated ones.

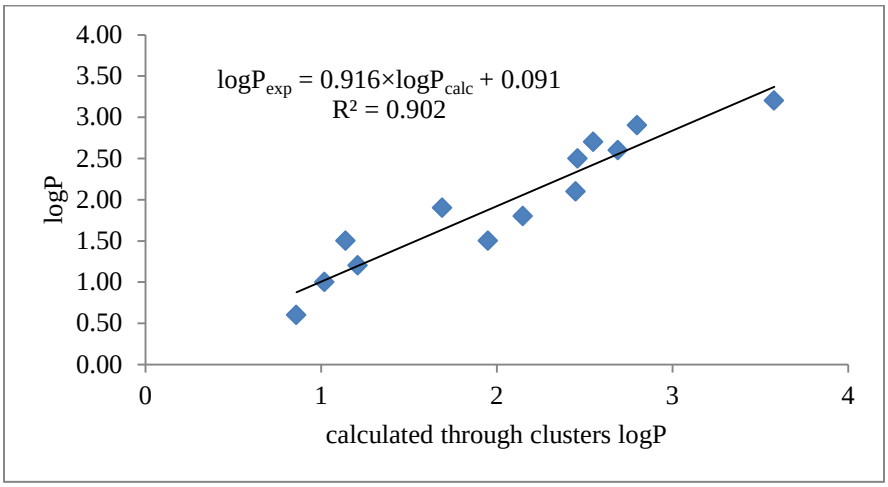


Fig.V.17. The plot $logP_{exp}$ vs. $logP_{calc.}$ for the test set (cluster similarity validation).

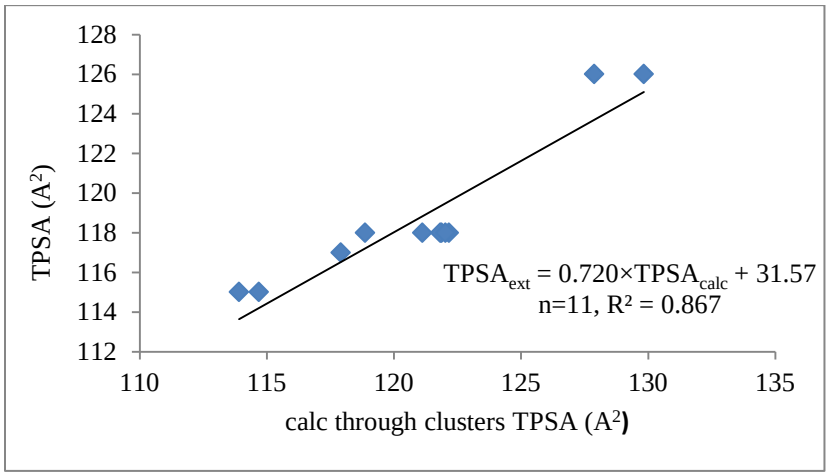


Fig.V.18. The plot $TPSA_{ext}$ vs. $TPSA_{calc.}$ for the test set (similarity cluster validation).

Prediction of logP (R^2=0.902), and TPSA (R^2= 0.867), is more accurate when using the similarity clusters, compared to the classical external validation of the model.

V.4. QSAR study on cephalosporines

The cephalosporins are a class of β-lactam antibiotics were discovered in 1945 by the Italian pharmacologist Giuseppe Brotzu and were first sold in 1964. Cephalosporins are bactericidal and have the same mode of action as other β-lactam antibiotics (penicillins), but are less susceptible to β-lactamases.

Table V.38. Downloaded drugs from cephalosporins family (https://pubchem.ncbi.nlm.nih.gov)

			logP			PSA			
Mol	Name	CID	Pubchem	AlogPS	Jchem	PubChem	Jchem	Polarizability (A^3)	Refractivity (A^3)
1	Cefacetrile	91562	-0.5	-0.52	-1.78	162	136.80	31.32	77.51
2	Cefaclor	51039	-1.8	0.85	-2.31	138	112.73	35.11	89.56
3	Cefadroxil	47965	-2.1	0.51	-2.45	158	132.96	35.86	90.95
4	Cefalotin	6024	-0.4	0.63	0.02	167	113.01	37.22	93.79
5	Cefamandole	456255	-0.9	-0.05	0.03	201	150.54	42.45	126.65
6	Cefapirin	30699	-1.1	0.18	-2.00	177	125.90	40.63	122.43
7	Cefazolin	33255	-0.4	-0.40	-1.52	235	156.09	41.44	119.86
8	Cefdinir	6915944	0	0.02	-1.69	212	158.21	36.12	94.34
9	Cefditoren	9870843	0.7	1.70	-0.15	242	160.10	49.19	124.18
10	Cefepime	5479537	-0.1	-0.37	-4.29	204	150.04	47.53	141.98
11	Cefixime	5362065	-0.7	0.25	-1.18	238	184.51	41.62	104.91
12	Cefmenoxime	9570757	0	-0.13	-0.83	270	190.81	47.04	133.51
13	Cefmetazole	42008	-0.6	-0.38	-0.65	239	163.33	44.50	124.57
14	Cefonicid	43594	-1.9	-0.71	-2.51	264	204.91	47.96	136.58
15	Cefoperazone	44187	-0.7	-0.11	-0.90	271	220.26	62.82	169.06
16	Ceforanide	43507	-3.2	-1.35	-3.17	244	193.63	49.27	139.87
17	Cefotaxime	5742673	-1.4	0.14	-1.41	227	173.51	41.77	105.11
18	Loracarbef	5284585	-1.7	0.55	-2.40	113	112.73	32.61	86.64
19	Cefotiam	43708	-2.4	-0.33	-3.09	251	172.46	49.87	142.34
20	Cefpiramide	636405	-0.1	0.53	0.22	259	212.76	58.79	164.81
21	Cefpodoxime	6335986	-1.4	0.05	-1.19	210	156.44	39.90	100.71
22	Cefprozil	5281006	-1.4	0.94	-1.92	158	132.96	39.35	101.27
23	Cefradine	38103	0.4	0.70	-2.45	138	112.73	33.19	92.00
24	Ceftazidime	5481173	0.4	-1.21	-4.12	245	191.22	51.06	143.88
25	Ceftizoxime	6533629	0	0.40	-0.85	201	147.21	35.38	89.90
26	Ceftobiprole	6918430	-3.7	-1.27	-4.53	250	203.44	50.41	131.04
27	Ceftolozane	53234134	-3.2	-1.20	-8.68	356	302.21	62.09	194.51
28	Ceftriaxone	5479530	-1.3	0.01	-1.79	288	208.98	51.47	128.47
29	Cefuroxime	5479529	-0.2	-0.24	-0.90	199	173.76	38.75	97.17
30	Cephalexin	27447	0.6	0.55	-2.14	138	112.73	32.52	89.97
31	Cephaloglycin	19150	-3	0.54	-2.90	164	139.03	37.64	99.90

Hypermolecule (Figure V.19.a-b) building was realized using some special features of NanoStudio software, “Connect” and “Identification”, atom by atom of each structure.

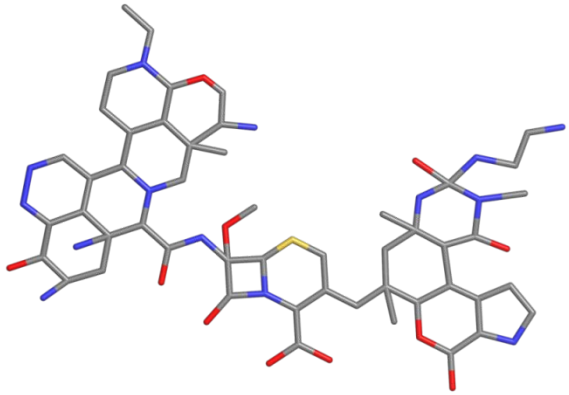


Fig.V.19.a. Hypermolecule, comprising the features of the dataset

Basic equation performed on entire set

$$logP = 16.415+0.998\times SD_{logP}; \quad n=27;\ R^2=0.865;\ s=0.435;\ F=160.164 \qquad (25)$$

QSAR Models

Mathematical models generated in training set are illustrated in Eq. 26-28.

Monovariate regression $logP = 16.924+1.0242\times SD_{logP}$ (26)

$n=18;\ R^2=0.874;\ s=0.452;\ F=111.735$

Bivariate regression $logP = 16.379+0.970\times SD_{logP}-14\times10^{-5}\times D3D$ (27)

$n=18;\ R^2=0.902;\ s=0.413;\ F=69.167$

Trivariate regression $logP = 14.109+0.960\times SD_{logP}-6\times10^{-4}\times D3D-3\times10^{-5}\times TE$ (28)

$n=18;\ R^2=0.923;\ s=0.380;\ F=55.810$

Model validation

Internal validation LOO

Table V.42. Leave-one-out results

D_1	D_2	D_3	R^2	Q^2	R^2- Q^2
SD_{logP}			0.875	0.845	0.030
SD_{logP}	D3D		0.902	0.866	0.036
SD_{logP}	D3D	TE/N	0.923	0.874	0.049

External validation of trivariate model on the test set, using Eq.28, conducts to result plotted in Figure V.20.

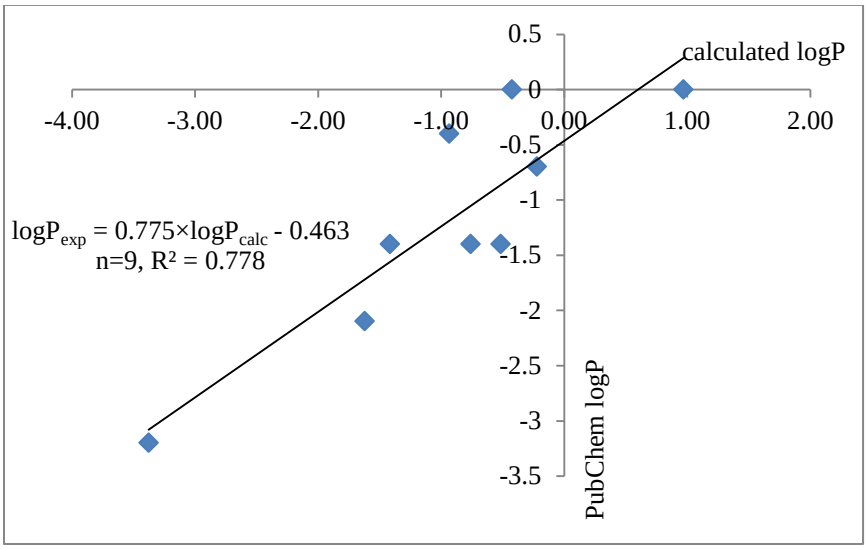


Fig. V.20. The plot logP$_{exp}$ vs. logP$_{calc.}$ for the test set (external validation).

5. QSAR study on penicillines

β-Lactam antibiotics inhibit the formation of peptidoglycan links in the bacterial cell wall; this is realized through binding of the four-membered *β*-lactam ring of penicillin to the enzyme DD-transpeptidase. It cannot catalyze formation of these cross-links, the imbalance between cell wall production and degradation causes the cell to die.

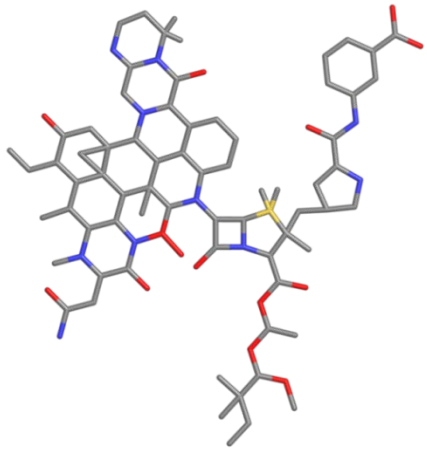


Fig. V.21.a Hypermolecule comprising the features of the dataset

Table V.44 lists 40 structures sharing the β-lactam ring.

Table V.44. penicillin derivatives with their CID and properties logP and LD_{50}.(https://pubchem.ncbi.nlm.nih.gov)

Mol	CID	Name	MW	XlogP3	LD50 rat, oral	Canonical SMILES
1	5493108	AC1NUTQ9	463.51	-2.9	2310.0	CC1(C(N2C(S1)C(C2=O)NC(C3=CC=CC=C3)C(=O)NC(=O)C(CC(=O)N)N)C(=O)O)C
2	5745669	AC1NX6WR	479.51	-3.7	4610.0	CC1(C(N2C(S1)C(C2=O)NC(C3=CC=C(C=C3)O)C(=O)NC(=O)C(CC(=O)N)N)C(=O)O)C
3	5745670	AC1NX6WT	477.54	-2.7	2280.0	CC1(C(N2C(S1)C(C2=O)NC(C3=CC=CC=C3)C(=O)NC(=O)C(CC(=O)N)NC)C(=O)O)C
4	71365	Almecillin PenO	330.42	1.8	~	CC1(C(N2C(S1)C(C2=O)NC(=O)CSCC=C)C(=O)O)C
5	36273	Amdinocillin	325.43	2.1	1544.3	CC1(C(N2C(S1)C(C2=O)N=CN3CCCCCC3)C(=O)O)C
6	33613	Amoxicillin	365.40	-2	1703.6	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=C(C=C3)O)N)C(=O)O)C
7	6249	Ampicillin	349.41	-1.1	1562.0	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=CC=C3)N)C(=O)O)C
8	71961	Aspoxicillin	493.54	-3.4	8000.0	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=C(C=C3)O)NC(=O)C(CC(=O)NC)N)C(=O)O)C
9	15574941	Azidocillin	375.40	2.8	2132.9	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=CC=C3)N=[N+]=[N-])C(=O)O)C
10	5284519	Azlocillin	461.49	0.1	2183.8	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=CC=C3)NC(=O)N4CCNC4=O)C(=O)O)C
11	441397	Bacampicillin	465.52	2.7	2073.3	CCOC(=O)OC(C)OC(=O)C1C(SC2N1C(=O)C2NC(=O)C(C3=CC=CC=C3)N)(C)C
12	6196	Oxacillin	401.44	2.4	1658.3	CC1=C(C(=NO1)C2=CC=CC=C2)C(=O)NC3C4N(C3=O)C(C(S4)(C)C)C(=O)O
13	20824	Carbenicillin	378.40	1.1	1459.9	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=CC=C3)C(=O)O)C(=O)O)C
14	33672	Carfecillin	454.50	3	980.0	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=CC=C3)C(=O)OC4=CC=CC=C4)C(=O)O)C
15	6098	Cloxacillin	435.88	2.4	1994.6	CC1=C(C(=NO1)C2=CC=CC=C2Cl)C(=O)NC3C4N(C3=O)C(C(S4)(C)C)C(=O)O
16	19003	Cyclacillin	341.43	1.3	1864.3	CC1(C(N2C(S1)C(C2=O)NC(=O)C3(CCCCC3)N)C(=O)O)C
17	18381	Dicloxacillin	470.33	2.9	1994.6	CC1=C(C(=NO1)C2=C(C=CC=C2Cl)Cl)C(=O)NC3C4N(C3=O)C(C(S4)(C)C)C(=O)O
18	71797	Ephicillin	433.57	2.9	~	CCN(CC)CCOC(=O)C1C(SC2N1C(=O)C2NC(=O)CC3=CC=CC=C3)(C)C
19	6438232	Flavicidin PenF	312.38	1.9	2000.0	CCC=CCC(=O)NC1C2N(C1=O)C(C(S2)(C)C)C(=O)O
20	12314049	Heptylpenicillin PenA	342.45	3.5	2000.0	CCCCCCCC(=O)NC1C2N(C1=O)C(C(S2)(C)C)C(=O)O
21	6087	Meticillin	380.42	1.2	1889.3	CC1(C(N2C(S1)C(C2=O)NC(=O)C3=C(C=CC=C3OC)OC)C(=O)O)C
22	36921	Ticarcillin	384.43	0.8	1459.1	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CSC=C3)C(=O)O)C(=O)O)C
23	656511	Mezlocillin	539.58	-0.2	2500.3	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=CC=C3)NC(=O)N4CCN(C4=O)S(=O)(=O)C)C(=O)O)C
24	8982	Nafcillin	414.48	2.9	2120.8	CCOC1=C(C2=CC=CC=C2C=C1)C(=O)NC3C4N(C3=O)C(C(S4)(C)C)C(=O)O
25	180562	O-Chlorophenoxymethyl penicillin	384.38	2.7	~	CC1(C(N2C(S1)C(C2=O)NC(=O)COC3=CC=CC=C3Cl)C(=O)O)C
26	6439405	Octenoylpenicillin	340.44	3	~	CCCCC=CCC(=O)NC1C2N(C1=O)C(C(S2)(C)C)C(=O)O
27	5852	Penicillamine	149.21	-1.8	2029.0	CC(C)(C(C(=O)O)N)S
28	115163	Pivmecillinam	439.57	3.1	2187.4	CC1(C(N2C(S1)C(C2=O)N=CN3CCCCCC3)C(=O)OCOC(=O)C(C)(C)C)C
29	167942	p-Hydroxy penicillin	366.39	1.7	~	CC1(C(N2C(S1)C(C2=O)NC(=O)COC3=CC=C(C=C3)O)C(=O)O)C
30	150610	Ertapenem	475.52	-1.5	2080.3	CC1C2C(C(=O)N2C(=C1SC3CC(NC3)C(=O)NC4=CC=CC(=C4)C(=O)O)C(=O)O)C(C)O
31	5904	Benzylpenicillin PenG	334.39	1.8	1652.3	CC1(C(N2C(S1)C(C2=O)NC(=O)CC3=CC=CC=C3)C(=O)O)C
32	123630	Tazobactam	300.29	-2	1810.0	CC1(C(N2C(S1(=O)=O)CC2=O)C(=O)O)CN3C=CN=N3
33	120720	Penicillin X (III)	350.39	1.5	1652.3	CC1(C(N2C(S1)C(C2=O)NC(=O)CC3=CC=C(C=C3)O)C(=O)O)C
34	71724	Adicillin PenN	359.40	-2.5	2000.0	CC1(C(N2C(S1)C(C2=O)NC(=O)CCCC(C(=O)O)N)C(=O)O)C
35	107556	Amylpenicillin PenDF	314.40	2.4	2000.0	CCCCCC(=O)NC1C2N(C1=O)C(C(S2)(C)C)C(=O)O
36	21319	Flucloxacillin	453.87	2.6	2083.4	CC1=C(C(=NO1)C2=C(C=CC=C2Cl)F)C(=O)NC3C4N(C3=O)C(C(S4)(C)C)C(=O)O
37	443387	Hetacillin	389.47	-0.6	1851.4	CC1(C(N2C(S1)C(C2=O)N3C(=O)C(NC3(C)C)C4=CC=CC=C4)C(=O)O)C
38	6869	Phenoxymethylpenicillin	350.39	2.1	1895.3	CC1(C(N2C(S1)C(C2=O)NC(=O)COC3=CC=CC=C3)C(=O)O)C
39	43672	Piperacillin	517.56	0.5	2392.5	CCN1CCN(C(=O)C1=O)C(=O)NC(C2=CC=CC=C2)C(=O)NC3C4N(C3=O)C(C(S4)(C)C)C(=O)O
40	33478	Pivampicillin	463.55	2.9	1919.3	CC1(C(N2C(S1)C(C2=O)NC(=O)C(C3=CC=CC=C3)N)C(=O)OCOC(=O)C(C)(C)C)C

There are 5 molecules with unknown biological activity. The purpose of this study is to predict their activities.

Average value of LD_{50} of entire set is 2167 mg/kg; this value is taken in consideration in SD_{LD50} determination.

Basic equation is obtained , molec 14 being an outlier.

$LD_{50}=1997.8+0.992\times SD_{LD50}$; n=39, $R^2=0.954$, s=233.751, F=761.981 (29)

The model used for prediction is by Eq.30 described.

$LD_{50}=1994.253+0.999\times SD_{LD50}$; n=34, $R^2=0.961$, s=230.279, F=791.147 (30)

Model validation

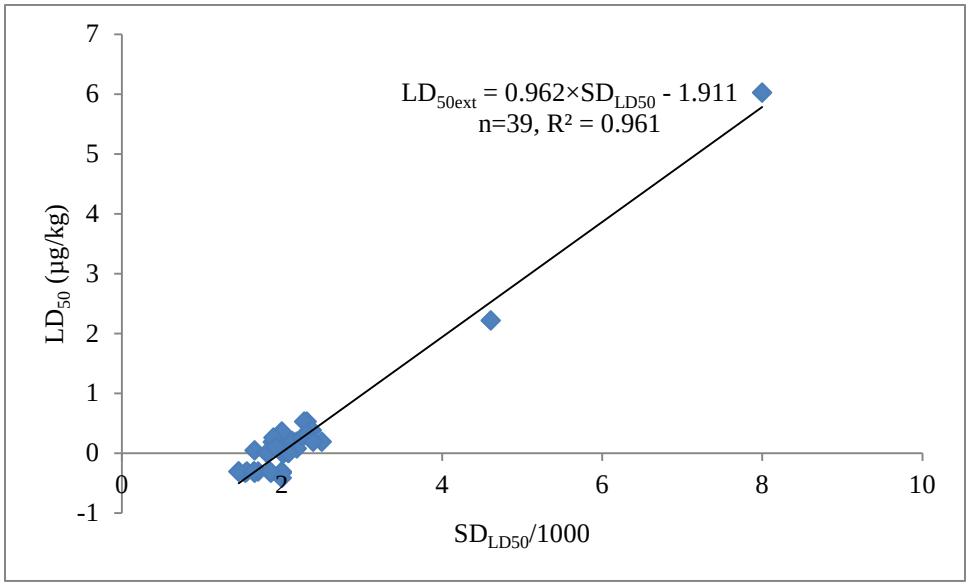


Fig. V.22. The plot $LD_{50}=f(SD_{LD50})$

The calculated values (predicted) of LD_{50} for those 5 molecules range successfully among the external values.

V.6. QSAR study on capsaicin derivatives

Capsaicin, $C_{18}H_{27}O_3N$, and several related compounds are called *capsaicinoids* and are produced as secondary metabolites by chili peppers.

The studied molecules downloaded from PubChem are listed in Table V.47.

Table V.47. Capsaicin derivatives with their CID, properties logP and TPSA.(https://pubchem.ncbi.nlm.nih.gov)

Mol	CID	Canonical SMILES	XlogP3	TPSA(A^2)
1	1548943	CC©C=CCCCCC(=O)NCC1=CC(=C(C=C1)O)OC	3.6	58.6
2	42759	CCCCCCCCNC(=O)C1=CC(=C(C=C1)O)OC	4.1	58.6
3	2998	CCCCCCCCC(=O)NCC1=CC(=C(C=C1)O)OC	4.2	58.6
4	3041745	CCCCCCCCNC(=O)CC1=CC(=C(C=C1)O)OC	4.2	58.6
5	69336196	CCCCCCOCCC(=O)NCC1=CC(=C(C=C1)O)OC	2.8	67.8
6	10090630	CC(=CCCC(=CCCC(=CC(=O)NCC1=CC(=C(C=C1)O)OC)C)C)C	5.8	58.6
7	25201160	CC(=CCCCCC(=O)NCC1=CC(=C(C=C1)O)OC)C	3.4	58.6
8	68413664	CCCCCCCCCCNC(=O)NCC1=CC(=C(C=C1)O)OC	5.0	70.6
9	3065261	CCCCCCCCCNC(=O)CC1=CC(=C(C=C1)O)OC	4.7	58.6
10	169252	CCCCCCCCCC(=O)NCC1=CC(=C(C=C1)O)OC	4.7	58.6
11	3053256	CCCCCCCCCCCC(=O)NCC1=CC(=C(C=C1)O)OC	5.8	58.6
12	160785	COC1=C(C=CC(=C1)CNC(=O)CCCCCCCCC=C)O	4.8	58.6
13	206278	CCCCCCCCCCC(=O)NCC1=CC(=C(C=C1)O)OC	5.3	58.6
14	3021470	CCCC(=O)NCC1=CC(=C(C=C1)O)OC	1.5	58.6
15	3022073	CCCCC(=O)NCC1=CC(=C(C=C1)O)OC	2.0	58.6
16	46887832	CC1CCC(C(C1)NC(=O)NCC2=CC(=C(C=C2)O)OC)C©C	4.0	70.6
17	68414474	COC1=C(C=CC(=C1)CNC(=O)NCC2=CC=CC=C2)O	1.3	70.6
18	3041816	CCCCCC(=O)NCC1=CC(=C(C=C1)O)OC	2.6	58.6
19	68760229	COC1=C(C=C(C=C1)CNC(=O)O)O	0.7	78.8
20	67419566	CC©©N(CC1=CC(=C(C=C1)O)OC)C(=O)O	1.9	70.0
21	22245803	CC©©OC(=O)NC1=CC(=C(C=C1)OC)O	2.2	67.8
22	5149140	CCCCCCCCC(=O)NCC1=CC(=C(C=C1I)O)OC	4.8	58.6
23	168836	CC©CCCCCC(=O)NCC1=CC(=C(C=C1)O)OC	3.9	58.6
24	9839519	CC©C=CCCCCC(=O)OCC1=CC(=C(C=C1)O)OC	4.2	55.8
25	101751387	CC(CCCCCCC(=O)NCC1=CC(=C(C=C1)O)OC)CO	3.0	78.8
26	20058472	CCCCCCCCNC(=O)C©C1=CC(=C(C=C1)O)OC	4.7	58.6
27	66552406	CCCCCCCCCNC(=O)C©C1=CC(=C(C=C1)O)OC	5.3	58.6
28	107982	CC©CCCCCCC(=O)NCC1=CC(=C(C=C1)O)OC	4.4	58.6
29	4446034	CCCCCCCCC(=O)NCC1=CC=CC=C1	4.6	29.1
30	44398654	CCCCCCCCC(=O)NCC1=CC(=C(C(=C1)OC)O)CC	5.0	58.6
31	66552464	CCCCCCCCC(=O)NC©C1=CC(=C(C=C1)O)OC	4.6	58.6
32	11630698	CC©CC=CCCC(=O)NCC1=CC(=C(C=C1)O)OC	3.1	58.6
33	11486920	CCCCCCCCC(=O)NCC1=C(C(=C(C=C1)O)OC)I	4.8	58.6
34	122189959	COC1=CC=C(C=C1)CC(=O)NCC2=CC(=C(C=C2)O)O	1.9	78.8
35	52944666	CC©©C1=CC=C(C=C1)CNC(=O)CC2=CC(=C(C=C2)O)OC	3.9	58.6
36	71541380	CN©CCNC(=O)CC1=CC(=C(C=C1)O)OC	0.8	61.8
37	15068906	CC1=CC(=CC=C1)CCNC(=O)CC2=CC(=C(C=C2)OC)OC	3.4	47.6
38	9566808	CC(=NNC(=O)CC1=CC(=C(C=C1)OC)OC)C2=CN=CC=C2	1.9	72.8
49	71748835	CC©(C=CCCCCC(=O)NCC1=CC(=C(C=C1)O)OC)O	2.2	78.8

40	2966952	CC©©OC(=O)NC1=CC(=C(C=C1)OC)Cl	3.2	47.6

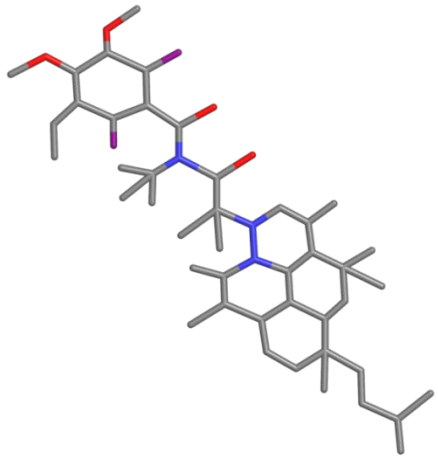


Fig.V.23.a. Hypermolecule comprising the features of the dataset

Basic equation (31) was performed on entire set

$logP = 4.377+SD_{logP}$; n=40; R^2=0.791; s=0.644; F=143.535 (31)

QSAR Models

Following already classic Diudea's algorithm, the QSAR model were established (Eq. 32-34)

Monovariate regression $logP = 4.381+0.992\times SD_{logP}$ (32)

n=28; R^2=0.815; s=0.672; F=114.300

Bivariate regression $logP = 5.045+1.012\times SD_{logP}-3.461\times Charges$ (33)

n=28; R^2=0.834; s=0.647; F=63.013

Trivariate regression $logP = 7.301+0.861\times SD_{logP}-2.856\times Charges+0.014\times TE/N$ (34)

n=28; R^2=0.846; s=0.637; F=43.997

Models validation

Internal validation LOO

Table V.51. LOO analysis results

D_1	D_2	D_3	R^2	Q^2	R^2- Q^2
SD_{logP}			0.815	0.791	0.024
SD_{logP}	Charges		0.834	0.798	0.037
SD_{logP}	Charges	TE/N	0.846	0.801	0.045

External validation of trivariate model, eq.34

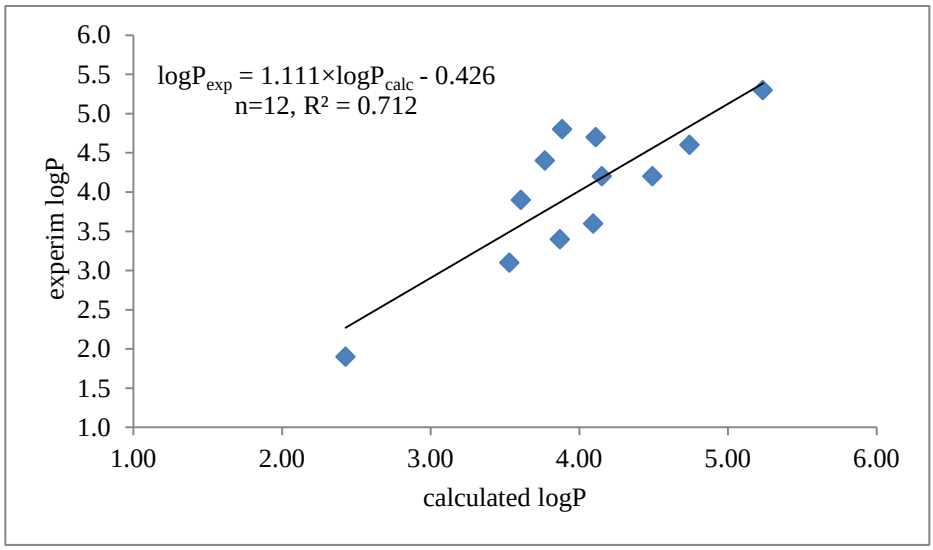


Fig.V.24. The plot logP$_{exp}$ vs. logP$_{calc.}$ for the test set (external validation).

Similarity Cluster Validation

The monovariate correlation for logP is plotted in Figure 25.

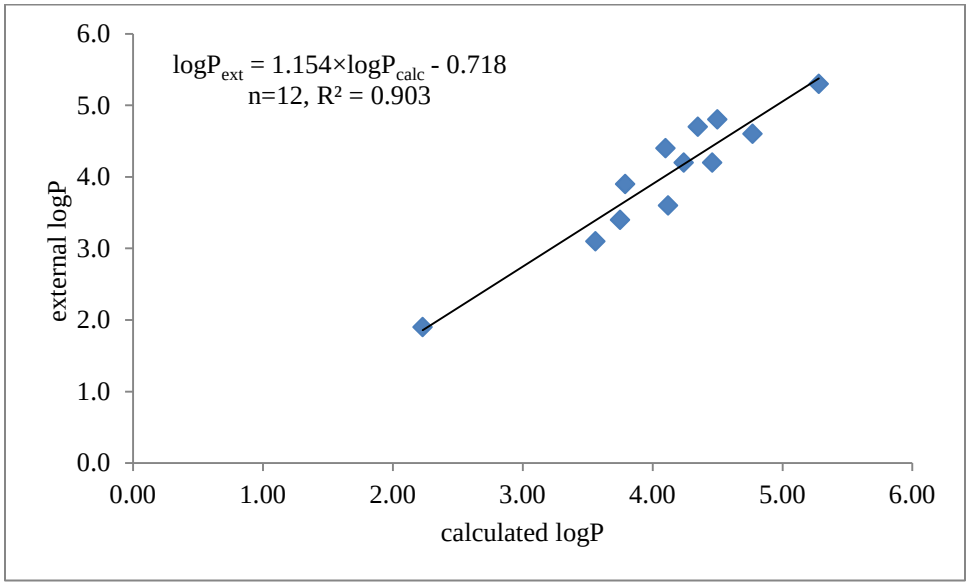


Fig.V.25. The plot logP$_{ext}$ vs. logP$_{calc.}$ for the test set (similarity cluster validation).

Prediction of logP (R^2=0.903) is more accurate when using the similarity clusters, compared to the classical external validation of the model (R^2=0.712).

V.7. QSAR conclusion

The paradigm of Corvin Hansch "similar molecules exhibit similar biological activity" is basic idea of QSAR studies. It does not represent a causalily relating structure and activity, but is simply a statistics.

We stress here that our goal was not to find the best models but to test an algorithm (developed at TOPO Group CLUJ), a methodology enabling an "ad-hoc" best prediction, within the "hypermolecule simulation of the biological receptor and regression weighting" of independent variables; the predictability of models in external sets is thus not necessary.

Predictions by similarity clustering is a special case of Leave-one-out procedure, where one molecule property is predicted by its similarity cluster computed among initial database molecules, this enabling the cluster to be congeneric, as initially asked by C. Hansch.

Good results we obtained on several sets of molecules taken from PubChem and other data bases.

We stress that our approach is not looking for predictability of QSAR equations in any data set. Our work is focused on ad-hoc best prediction on a study case.

Prediction could be made for molecules that have no external activity taking the mean value of the set. Some descriptor can be calculated and then their activity predicted, some exemples been given in this thesis.

Six sets of bioactive organic molecules were studied and presented in this chapter:

1. Phenothiazines
2. Sumatriptans
3. Sildenafils
4. Cephalosporins
5. Penicilline
6. Capsaicines

VI. Final Conclusions

VI.1. General Conclusions

Within this Thesis,

1. Design of various multi-shell clusters of icosahedral and octahedral symmetry was accomplished by map operations and their structure described by enumerating the substructures (*i.e.*, figure counting), which enabled to establish their rank (3D or higher).
2. Equivalence classes of figures (vertices/atoms, edges/bonds, facets of higher rank) in icosahedral, octahedral or rhombohedral clusters have been solved by using the centrality topological index, which provides an image of atom distribution according to molecular centrality. Classes were confirmed by group symmetry calculations obtained by permutations in the adjacency matrix.

   Equivalence class counting is related to "topological symmetry", that is the maximum possible symmetry achievable by a molecular structure.

3. Energy of graphs was computed from the eigenvalues in propellanes (a new class of rhomboidal polytopes) and applied in a QSPR study performed on C_{40} fullerenes.
4. QSAR studies were performed in 6 classes of organic compounds with biological activity: phenothiazines, triptans, sildenafil derivatives, cephalosporines, penicillines, capsaicin derivatives.
5. Research resulted in 8 published articles, a chapter in a Springer book, 3 sent articles, 6 oral conference presentations, 3 conference posters

References (Selection)

Bhattacharya D, Klein DJ, Ortiz Y (2016) The astounding buckyball buckyball. *Chem Phys Lett* 647:185-188

Blatov VA, O'Keeffe M, Proserpio DM (2010) Vertex-, face-, point-, Schläfli-, and Delaney-symbols in nets, polyhedra and tilings: recommended terminology, CrystEngComm. 12, 44–48.

Ciubotariu D (1987) Structură-reactivitate în clasa derivaţilor acidului carbonic, PhD Thesis, Timisoara, Romania

Coxeter HSM (1974) Regular complex polytopes. Cambridge Univ. Press, Cambridge; Section 11.3 Petrie polygons.

Dinca M, Ciger S, Stefu M, Gherman F, Miklos K, Nagy CsL, Ursu O, Diudea MV (2004) Stability prediction in small fullerenes. *Carpath J Math 20*(2): 211-221

Diudea MV (1979) "*Fenotiazine și medicamente structural înrudite*", PhD Thesis, Inst. Chem. Cluj, 1979.

Diudea MV (2001), *QSPR/QSAR "Studies by Molecular Descriptors"*, NOVA, New York.

Diudea MV (2004) Covering forms in nanostructures. *Forma* (Tokyo) 19:131–163

Diudea MV (2010) *Nanomolecules and Nanostructures – Polynomials and Indices*, MCM, No 10, Univ Kragujevac, Serbia

Diudea MV (2013*) Quasicrystals: between spongy and full space filling*. In: Diudea MV, Nagy CL (Eds.) Diamond and Related Nanostructures, Springer, Dordrecht, pp. 335–385

Diudea MV (2015) 4D-Octahedral structures. *Int J Chem Model* (accepted)

Diudea MV (2016) *Multi-shell polyhedral clusters*, Springer (in preparation).

Diudea MV, Gutman I, Jäntschi L (2002) *Molecular Topology*, NOVA, New York.

Diudea MV, Nagy CL (2007) *Periodic nanostructures*, Springer, Dordrecht

Diudea MV, O. Ursu (2003) Layer matrices and distance property descriptors. *Indian J Chem* 42A:1283-1294

Diudea MV, Pârv B, Ursu O (2003) TORUS, Univ Babes-Bolyai, Cluj

Diudea MV, Pîrvan-Moldovan A, Kooperazan-Moftakhar F, Ashrafi AR (2016) Topological symmetry of multi-shell clusters. In: A. R. Ashrafi, M. V. Diudea, Eds. *Distance, Symmetry, and Topology in Carbon Nanomaterials*, Springer Int. Pub. Switzerland, pp. 61-82.

Diudea MV, Pîrvan-Moldovan A, Kooperazan-Moftakhar F, Ghorbani M (2016), Topological symmetry of high rank and genus clusters, *Symmetry: Culture and Science* (Budapest). 2016 (accepted).

Diudea MV, Pîrvan-Moldovan A, Pop A, Medeleanu M(2017) Energy of Graphs and Remote Graphs in Hypercubes, Propellanes and Fullerenes. *MATCH Commun. Math. Comput. Chem.* (accepted*)*.

Diudea MV, Rosenfeld VR (2017) The truncation of a cage graph, *J. Math. Chem*. *55*(4):1014-1020.

Diudea MV, Stefu M, John PE, Graovac A (2006) Generalized operations on maps, *Croat Chem Acta* 79:355–362

Euler L (1758) Elementa doctrinae solidorum. Novi Comm Acad Scient Imp Petrop 4:109–160

Fathalikhani K, Pîrvan-Moldovan A, Diudea MV (2016) Topological indices in hyper-tubes of hypercubes, *Studia Univ. Babes-Bolyai, Chemia*, *61*(1), 291-304.

Florkowski SF (2008-2012) Spectral graph theory of the hypercube; thesis, Naval postgraduate school, Monterey, California

Fowler PW (1986) How unusual is C_{60}? Magic numbers for carbon clusters. *Chem Phys Lett* 131:444–450

Free SM, Wilson JW (1964) A Mathematical Contribution to Structure-Activity Studies, *J. Med. Chem.*, *7*, 395.

Furusjö E, Svenson A, Rahmberg M, Andersson M (2006) The importance of outlier detection and training set selection for reliable environmental QSAR predictions. *Chemosphere*; 63:99-108.

Geladi P, Kowalski BR (1986) Partial least squares regression: a tutorial. *Anal Chim Acta* 185:1-17.

Goldberg M (1979) Convex polyhedral space-fillers of more than twelve faces. *Geom Dedicata* 8: 491-500

Grünbaum B (2003) *Convex Polytopes*, Graduate Texts in Mathematics (2nd ed.), Springer.

Hafner I, Zitko T (2002a) Introduction to golden rhombic polyhedra. *Visual Mathematics* 4(2):2(3)

Hafner I, Zitko T (2002b) Relations among rhombic, Platonic and Archimedean solids. *Visual Mathematics* 4(2):2(4)

Hansch C, Leo A (1985) *Exploring QSAR*. Washington, DC: American Chemical Society.

Jäntschi L (2005) LOO Analysis (LOO: leave one out), AcademicDirect Library of software, Available at: http://l.academicdirect.org/Chemistry/SARs/SARs/loo/

Johnson NW (1966) *The Theory of Uniform Polytopes and Honeycombs*, Ph. D. Dissertation, University of Toronto.

Katz MH (2006) *Multivariable Analysis – A Practical Guide for Clinicians.* 2nd Edition. Cambridge University Press.

Kooperazan-Moftakhar F, Pârvan-Moldovan A, Diudea MV (2015) Topological symmetry of multi-shell octahedral clusters. *Int J Chem Model* (accepted).

Kubinyi H (1993) "*QSAR: Hansch analysis and related approaches*", R. Mannhold, Krogsgaard-Larsen P., Timmerman H. (Eds.),VCH Publishers, New York, 1993.

McMullen P, Schulte E (2002) *Abstract Regular Polytopes*, 1st ed., Cambridge Univ. Press, Cambridge.

Moldovan DC, Costescu A, Katona G, Diudea MV (2008) *MATCH Commun*. *Math*. *Comput*. *Chem.*, *60*, 977.

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. J.Comput.Chem. (16):2785-91.

Nagy CL, Diudea MV (2005) JSChem, Univ Babes–Bolyai, Cluj

Nagy CL, Diudea MV (2009) Nano Studio sofrware program. Babes-Bolyai University, Cluj.

Nagy CL, Diudea MV (2017) Ring signature index. *MATCH Commun Math Comput Chem* 77(2): 479-492.

Pârvan-Moldovan A, Diudea MV (2015a) Cell@cell higher dimensional structures. *Studia Univ Babes-Bolyai", Chemia* 60 (2):379–388.

Pârvan-Moldovan A, Diudea MV (2015b) Hyper-tubes of hyper-cubes, *Iran. J. Math. Chem., 6*, 163-168.

Pârvan-Moldovan A, Kooperazan-Moftakhar F, Diudea MV (2014) Topological symmetry of multi-shell icosahedral clusters. *Studia Univ "Babes-Bolyai", Chemia* 59(3):103–108.

Pașca RD, Moldovan A, Horovitz O, Lupan A (2017) Thermodynamic investigation of tautomerism in triazine oximes, *JOMC* (sent).

Pîrvan-Moldovan A, Diudea MV (2016) Euler's characteristic in polyhedral graphs transforming, *Croat Chem. Acta*, 89(4) DOI: 10.5562/cca300.

Pîrvan-Moldovan A, Ersali S (2016) QSAR study on phenothiazines, *Studia Univ. Babes-Bolyai", Chemia, 61* (1), 305-315.

Randić M (1990) The nature of chemical structure. *J Math Chem*, 4:157-84.

Randić M (1993) Novel molecular description for structure-property studies. *Chem Phys Lett* 211:478-483

Randić M, Razinger M (1997) On characterization of 3D molecular structure. In: Balaban AT (ed.*): From chemical topology to three dimensional geometry*. New York: Plenum Press.

Schulte E (1985) Regular incidence-polytopes with Euclidean or toroidal faces and vertex-figures, *J. Combin. Theory, Series A* 40 305–330.

Stefu M, Diudea MV (2005) CageVersatile_CVNET software program, Babeș-Bolyai University, Cluj.

Stefu M, Pârvan-Moldovan A, Kooperazan-Moftakhar F, Diudea MV (2015) Topological symmetry of C_{60}-related multi-shell clusters. *MATCH Commun. Math. Comput. Chem., 74*, 273-284.

The GAP Team, GAP – Groups, Algorithms, Programming - a System for Computational Discrete Algebra, GAP 4.7.5 release. http://www.gap-system.org.

Topliss JG, Costello RJ (1972) Change correlations in structure-activity studies using multiple regression analysis, *J. Med. Chem.***,** *15*, 1066.

Toropov AA, Toropova AP (2001) Modeling of lipophilicity by means of correlation weighting of local graph invariants, *J. Mol. Struct.* (*Theochem*), *538*, 197.

Ursu O, Diudea MV (2005) TOPOCLUJ software program, Babes-Bolyai University, Cluj.

Wald A (1939) Contributions to the Theory of Statistical Estimation and Testing Hypotheses, *Ann. Math. Statist.* 10, *4*, 299-326.

Wiberg KB, Walker FH (1982) [1.1.1]Propellane, *J. Am. Chem. Soc.*, 1982, 104 (19), 5239–5240.

Willett P, Barnard JM, Downs GM (1998) Chemical Similarity Searching *J. Chem. Inf. Comput. Sci., 38(6),* 983.