Materials Discovery 5 (2016) 14–21

Contents lists available at ScienceDirect

Materials Discovery

journal homepage: www.elsevier.com/locate/md

Elemental factorial study on one-cage pentagonal face nanostructure congeners

Lorentz Jäntschi [a,b,*], Donatella Bálint [b], Lavinia L. Pruteanu [b], Sorana D. Bolboacă [c]

[a] Department of Physics and Chemistry, Technical University of Cluj-Napoca, Cluj-Napoca, Romania
[b] Chemistry Doctoral School, Babeş-Bolyai University, Cluj-Napoca, Romania
[c] Department of Medical Informatics and Biostatistics, Iuliu Haţieganu University of Medicine and Pharmacy, Cluj-Napoca, Romania

ARTICLE INFO

Article history:
Received 12 August 2016
Received in revised form 6 December 2016
Accepted 7 December 2016
Available online 8 December 2016

Keywords:
Factorial analysis
Nanostructure
Structure-property relationships

ABSTRACT

A full factorial design with two factors applied on the family of 81 dodecahedrane congeners is presented. One of the factors used four levels (the layers of the structure), while the other factor used three levels (the atom as Boron, Carbon, or Nitrogen, with the same atom for the layer). Ten calculated properties were input for investigation of the link between properties and structural features. Boron, Carbon or Nitrogen were considered as a reference atom. The models with determination coefficient near 1 comprised 22 to 44 distinct factors. The complexity of the models increases from boron taken as a reference to carbon taken as reference. Therefore, along with the less complexity with the factorial analysis (here elements were accounted in a trinomial scale), alternatives for the reference should also be sought (the available software packages for this type of regression do not check when the table of transformation from multinomial data type to binomial variables is built).

1. Introduction

Dodecahedrane ($C_{20}H_{20}$), a chemical spherical compound with high symmetry, has in each vertex a carbon atom that bonds three neighboring carbon atoms and one hydrogen atom [1]. The compound was first synthesized at the Ohio State University in 1982 [1–3] but theoretical studies in regards of group theory, graph theory, or molecular orbital theory were conducted since 1975 [4–7]. The dodecahedrane is a Platonic structure [8] and the compound and/or its derivatives have been investigated in theoretical [9–12] and experimental [13–16] settings. Substitution of all hydrogen atoms in $C_{20}H_{20}$ led to development of several compounds such as $C_{20}F_{20}$ (perfluorododecahedranes), $C_{20}Cl_{20}$, $C_{20}(OH)_{20}$ or $C_{20}Br_{20}$ [17]. Perfluorododecahedranes, which initially was reported as unstable [17], was successfully produced by Banfalvi in milligram quantities [18]. Dodecahedrane complexes were also of scientific interest as endo- ($X@C_{20}H_{20}$, where X = H, He, Ne, Ar, Li, Li^+, Be, Be^+, Be^{2+}, Na, Na^+, Mg, Mg^+, and Mg^{2+} [19] or X = C^{4-}, N^{3-}, O^{2-}, F^-, Ne [20]) or exo-hedral ($XC_{20}H_{20}$, X = H^+, H, N, P, C^-, Si^-, O^+, S^+ [21]) complexes. Dodecahedrane is of scientific interest due to its stability given by its high symmetry, covalent bond angles (108 °C) closest to the ideal tetrahedral bond angle (109.5°) and its C—C bound close to that of diamond (1.54 Å) [20]. Substitution of hydrogen atoms in dodecahedrane have been reported in the scientific literature [17,18,2020] but no study has investigated the substitution of carbon atoms. Furthermore, if the carbon atoms are substituted by other atom(s), the dodecahedrane become a congener [22]. An analysis of the link between the type of the atom and the properties in dodecahedrane may provide useful information on the diversity and stability of these structures. A theoretical study was conducted with the full factorial design to investigate the effects of factors and/or their interactions on ten calculated properties using the full class of dodecahedrane congeners when the molecule is seen as a four layered structure, as well as to investigate the effect on each layer when formed by carbon, boron or nitrogen.

2. Methodology

2.1. One-cage pentagonal faces nanostructures

The elementary unit of one-cage pentagonal faces structure (Fig. 1) was used to construct the full class of compounds using three atoms from the same period to assure the formation of at least three bonds (Boron, Carbon, and Nitrogen). The elementary structure was seen as a four layer structure (see Fig. 1) with each layer being occupied by one type of atom (Boron, Carbon or Nitrogen).

The full class of congeners obtained by this method was constructed based on the one-cage pentagonal faces structure

* Corresponding author at: Department of Physics and Chemistry, Technical University of Cluj-Napoca, Cluj-Napoca, Romania.
E-mail addresses: lorentz.jantschi@gmail.com, lori@academicdirect.org (L. Jäntschi), sbolboaca@umfcluj.ro (S.D. Bolboacă).

http://dx.doi.org/10.1016/j.md.2016.12.001
2352-9245/

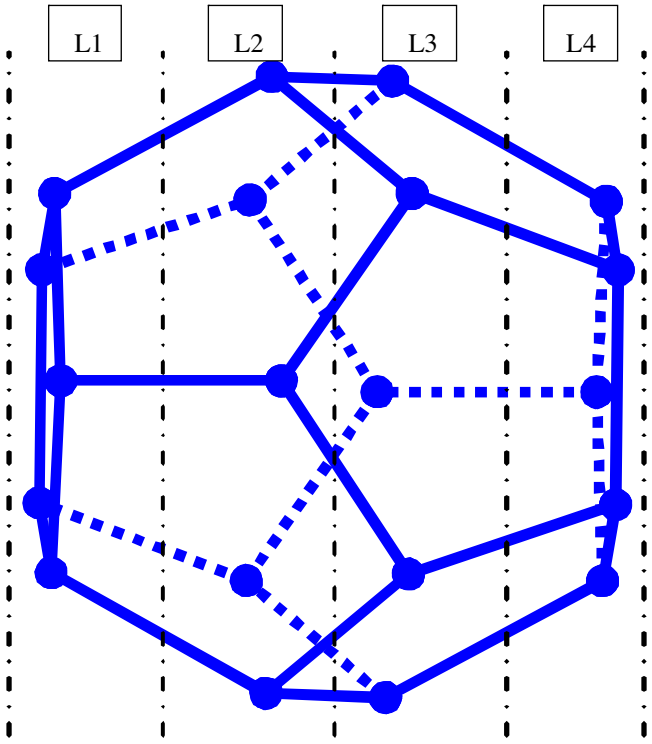


Fig. 1. Layers (L) on one-cage pentagonal faces structure.

presented in Fig. 1 and comprises 81 structures ($3^4 = 81$, where 3 is the type of the atom and 4 are the number of layers). Due to the symmetry and freedom of movement and rotation, just 45 structures are distinct (see Table 1).

The distinct structures of this family of nano-cages were study using full-factorial analysis design. The flowchart of the applied approach is depicted in Fig. 2. The cages were initially drawn with HyperChem software (geometry was initially optimized with PM3; Hypercube Inc. USA). The geometry of the cages was refined at a Hartree-Fock (RHF) level of theory [23–25] with 3–21G basis set, and further refined by Möller-Plesset [26] untill the second order (MP2) with 6–31G* basis set [27] using Spartan software (v.10, Wavefunction Inc. USA). The properties included in the full factorial analysis were calculated after geometry optimization (MP2 based calculations).

2.2. Full-factorial analysis on dodecahedrane congeners

The main issue in the proposed approach is related with the management of a large system of equations that must be solved (for one cage, with disregards of the symmetry, the system square matrix contains 729 variables). The implemented approach operates with the type of the atom (a quality) and accordingly the multinomial problem was translated into a multi-binomial problem [28]. As example, 'atom type' (as Boron, Carbon or Nitrogen) on a certain layer in Fig. 1 is a trinomial variable. The presence of a certain atom (e.g. B) was encoded with 1 (other atoms with 0), and thus the problem is translated from trinomial to tri-binomial when all three atoms are considered. One atom was taken as a refer-

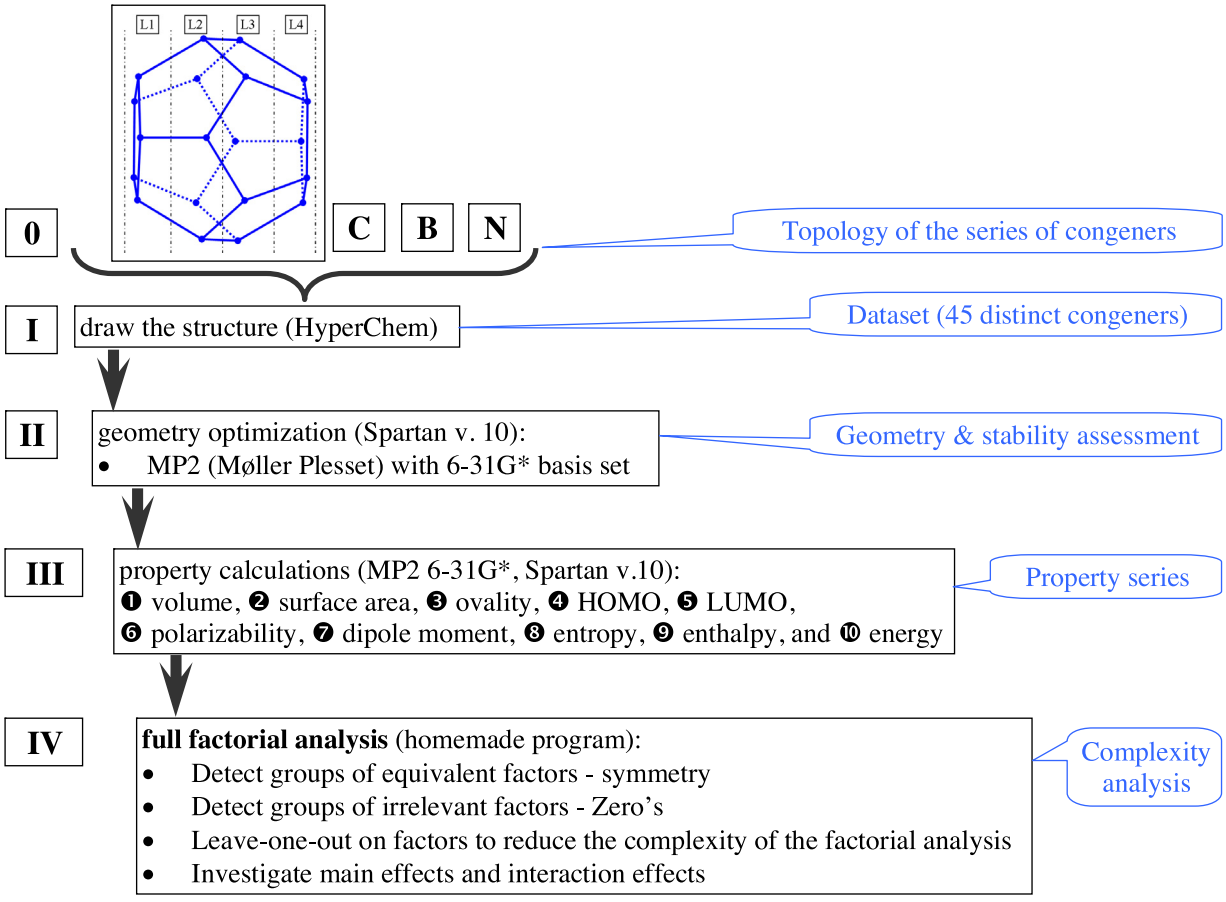


Fig. 2. Full factorial analysis flowchart on one-cage pentagonal faces structures with the same atom per layer (Boron or Carbon or Nitrogen). HOMO stands for highest occupied molecular orbital and LUMO is the lowest unoccupied molecular orbital.

ence, and the remaining ones were encoded as binary variables. The analysis was conducted by taking each atom (B, C or N) as reference atoms resulting in three distinct scenarios.

The algorithm presented in the box bellow was implemented as an executable program to solve the proposed full-factorial design:

```
For each different encoding letter (for each atom type)
    Take one letter as reference;
    Transform the input space into factors space using the referential;
    Conduct the full factorial analysis (calculate the coefficients of
    the factors);
    Write the result in the output file;
    Calculate machine precision (result: 1.08…×10^-19);
    Calculate eps as cubic root of the machine precision (result:
        4.76…×10^-7) and its number of significant digits (result: 7);
    With the eps detect the irrelevant factors (having the coefficient
        in the absolute value less than eps) and with the number of
        significant digits find the symmetry groups (having the same
        value of the coefficient for the digits till the number of
        significant digits);
    Remove the irrelevant factors from the model equation and group the
        symmetry factors in the model equation; conduct the analysis
        (obtain the coefficients of the orthogonal design);
    Write the result in the output file;
    Repeat
        Set the factor to be removed to first one;
        For each factor and factor group
          Make a copy of the model with the factor removed;
          Calculate the coefficients, their significances (Student t-
              values) and the significance of the model (Fisher's F-
              value);
          If the obtained F-value is greater than the value for the
              factor to be remove, then set the factor to be removed;
        EndFor;
        Remove the factor to be removed (with the highest F-value);
        Write the result in the output file (coefficients and significances);
    Until the number of factors and factor groups become zero.
EndFor
```

The first model in the output file of the implemented algorithm is the full factorial model, which embeds the influence of all factors and all interactions on the investigated system. Feeding the algorithm with all possible data opens the possibility to detect the symmetry (factors that have exactly the same effect are grouped, providing a symmetry group) and to identify the irrelevant factors (factors having a contribution with four or more orders of magnitude less than other factors). The implemented algorithm removes in each step the factor (or the factor group(s)) that after removal led to the highest statistical significance (highest F-value). This approach reduces the complexity of the models by keeping the most relevant factors and interactions.

Full factorial analysis was conducted for ten properties (from MP2 based calculations, see Fig. 2; entropy and enthalpy are for standard conditions, e.g. S^0 and H^0) with three models for each property (one model for Boron as the reference atom, one model for Carbon as the reference atom, and another model for Nitrogen as the reference atom).

3. Results and discussion

3.1. Distinct factors

The number of distinct factors in the orthogonal array, the number of symmetry constrains factors (factors with identical value) and the number of almost null coefficients for each investigated reference atom (B, C or N) are presented in Table 2. Almost null coefficient means that the value is outside of the precision for floating point calculations, 14 significant digits (e.g. modulus lower than 10^{-14}). Note that the number of factors in the orthogonal array is closed to the number of distinct congeners when the analysis is conducted on the whole family of dodecahedrane one-cage nanostructures that comprise 81 compounds.

An analysis of Table 2 revels that the starting point in regards to distinct number of factors and symmetry varied from property to property and depends on the reference atom. However, no differences are observed in regards of the distinct number of factors and symmetry constrains factors for three properties (dipole moment, entropy, and enthalpy) regardless of the reference atom. Almost null coefficients were identified in two cases, represented by ovality (8) and HOMO (highest occupied molecular orbital) (1) when C is the reference atom.

3.2. Evolution pattern of the correlation coefficient

The implemented algorithm identifies in every step the factor that by removal leads to the model with highest F-value. The pattern of correlation coefficient according to the decreasing number of factors for each investigated property is presented in Fig. 3.

The analysis of the trends presented in Fig. 3 reveals similar evolution of correlation coefficient for the same property when B and respectively N is the reference atom at least until number of factors in the equation is equal to four. When excluding enthalpy, the carbon as the reference atom led to slightly better performances, expressed as slightly higher values of correlation coefficients.

The relationships with a high number of factors (for volume: 44 factors when B is the reference atom, 43 factors when N is the

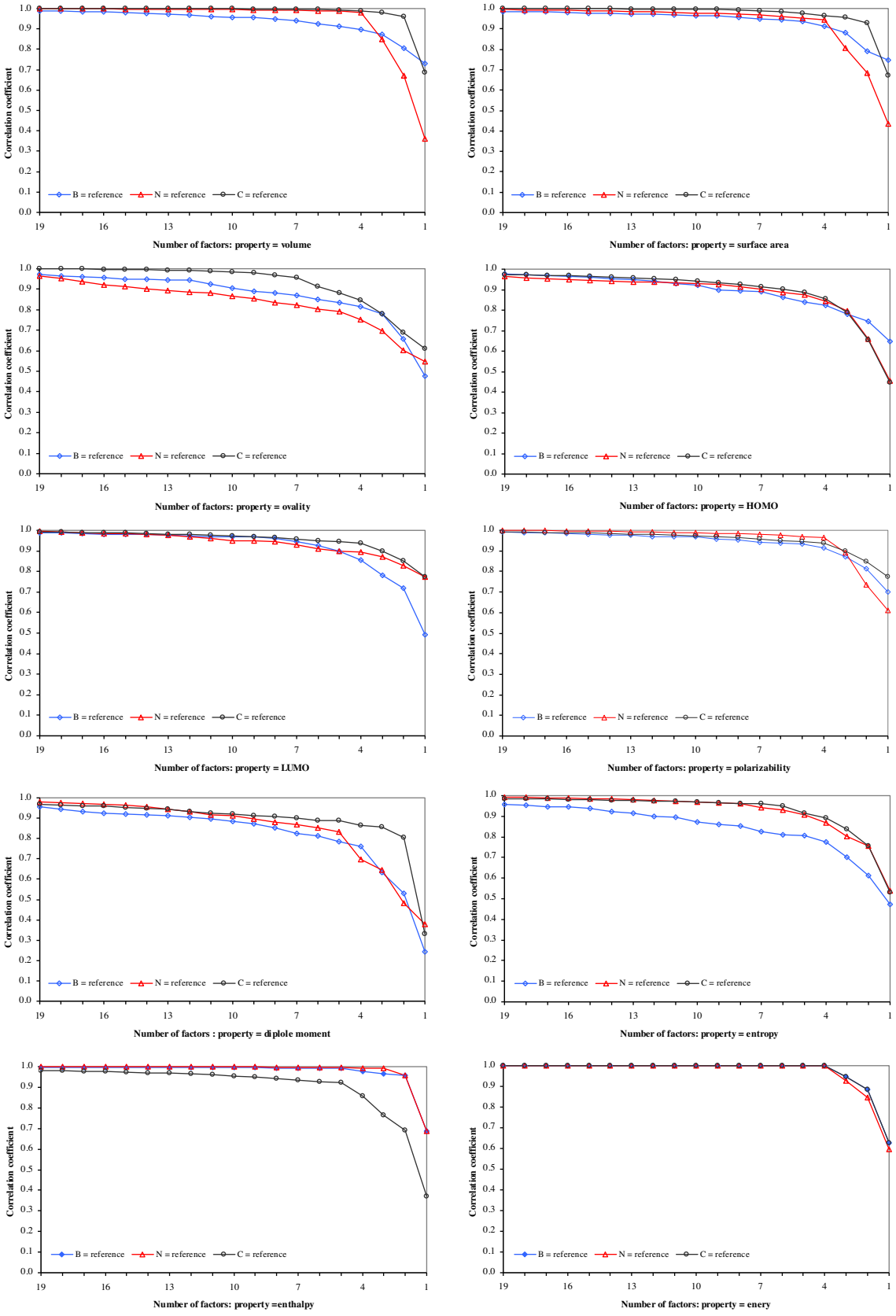


Fig. 3. Evolution of correlation coefficient with the decrease of number of factors in the explanatory equation: volume, surface area, ovality, HOMO (highest occupied molecular orbital), LUMO (lowest unoccupied molecular orbital), polarizability, dipole moment, and entropy. Evolution of correlation coefficient with the decrease of number of factors in the explanatory equation: enthalpy and energy.

Table 1
Full class of distinct congeners of one-cage pentagonal faces nanostructure.

ID	L1	L2	L3	L4	ID	L1	L2	L3	L4	ID	L1	L2	L3	L4
01	B	B	B	B	16	C	B	B	B	31	C	C	N	C
02	B	B	B	N	17	C	B	B	C	32	C	C	N	N
03	B	B	C	N	18	C	B	B	N	33	C	N	B	B
04	B	B	N	B	19	C	B	C	B	34	C	N	B	N
05	B	B	N	N	20	C	B	C	N	35	C	N	C	B
06	B	C	B	B	21	C	B	N	B	36	C	N	C	N
07	B	C	B	N	22	C	B	N	C	37	C	N	N	B
08	B	C	C	B	23	C	B	N	N	38	C	N	N	C
09	B	C	C	N	24	C	C	B	B	39	C	N	N	N
10	B	C	N	B	25	C	C	B	C	40	N	B	B	N
11	B	C	N	N	26	C	C	B	N	41	N	B	N	N
12	B	N	B	N	27	C	C	C	B	42	N	C	B	N
13	B	N	C	N	28	C	C	C	C	43	N	C	C	N
14	B	N	N	B	29	C	C	C	N	44	N	C	N	N
15	B	N	N	N	30	C	C	N	B	45	N	N	N	N
B = Boron					C=Carbon					N=Nitrogen				

Table 2
Summary of distinct factors (Dist), symmetry constrains factors (Sym), and almost null coefficients (∼Zero) according with property and reference atom.

Property_RefAtom	Dist	Sym	∼Zero	Property_RefAtom	Dist	Sym	∼Zero
Volume_B	44	36		Polarizability_B	43	37	
Volume_C	41	39		Polarizability_C	44	36	
Volume_N	43	37		Polarizability_N	44	36	
SurfaceArea_B	44	36		DipoleT_B	44	36	
SurfaceArea_C	43	37		DipoleT_C	44	36	
SurfaceArea_N	44	36		DipoleT_N	44	36	
Ovality_B	44	36		Entropy_B	44	36	
Ovality_C	22	50	8	Entropy_C	44	36	
Ovality_N	41	39		Entropy_N	44	36	
HOMO_B	44	36		Enthalpy_B	44	36	
HOMO_C	42	37	1	Enthalpy_C	44	36	
HOMO_N	44	36		Enthalpy_N	44	36	
LUMO_B	44	36		Energy_B	44	36	
LUMO_C	44	36		Energy_C	44	36	
LUMO_N	43	37		Energy_N	43	37	

Table 3
Relationships explaining no less than 95% of the change in the property of interest.

Property	RefAtom	NoFactors	F-value	R	Property	RefAtom	NoFactors	F-value	R
Volume	B	9	81	0.9548	Polarizability	B	8	88	0.9526
	C	2	450	0.9593		C	3	314	0.9615
	N	4	484	0.9809		N	4	269	0.9665
Surface area	B	8	97	0.9566	Dipole moment	B	19	35	0.9566
	C	3	277	0.9567		C	15	42	0.9526
	N	5	150	0.9536		N	14	49	0.9549
Ovality	B	15	40	0.9502	Entropy (S^0)	B	18	34	0.9533
	C	7	116	0.9578		C	7	123	0.9601
	N	18	32	0.9508		N	8	112	0.9622
HOMO	B	14	48	0.9540	Enthalpy (H^0)	B	2	416	0.9562
	C	12	57	0.9535		C	9	74	0.9504
	N	17	36	0.9523		N	2	416	0.9562
LUMO	B	8	103	0.9591	Energy	B	4	$5.83{\cdot}10^6$	0.9999
	C	7	114	0.9573		C	4	$5.83{\cdot}10^6$	0.9999
	N	10	65	0.9505		N	4	$5.83{\cdot}10^6$	0.9999

RefAtom = reference atom (B = Boron, C=Carbon, N=Nitrogen).
NoFactors = number of factors in the model; F-value = F statistic of the model; R = correlation coefficient.

reference atom, and 41 when C is the reference atom, see Table 2) assure a model with correlation coefficient closest to one. However, a correlation coefficient closest to one is not a sufficient criterion in assessment of a model [29–32].

It is expected that the full relationships to be provided by the same number of factors (the expected number of different factor groups is equal with the number of distinct structures in the family − 45, see Supplementary Table) but in fact different number of factors for different atom as reference atom is observed. Furthermore, even if the number of factors is not different, differences in the group of factors for different reference atoms are observed (see Supplementary Table). However, the number of different factor groups is actually 44 + 1 only when Boron ('B', Supplementary Table) is used as reference atom, is 41 + 1 when Carbon is the reference atom ('C', Supplementary Table) and is 43 + 1 when Nitrogen ('N', Supplementary Table) is the reference atom. Therefore, we

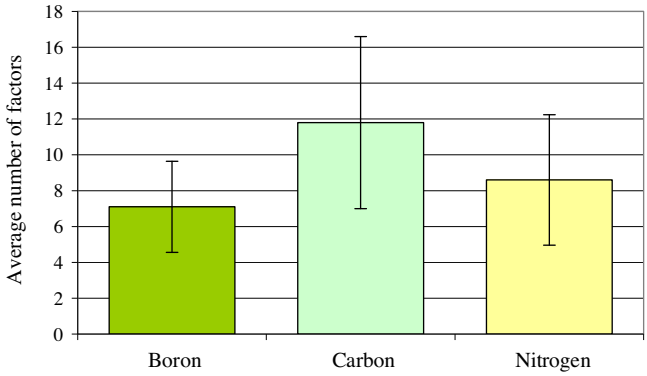


Fig. 4. Average number of factors in the first models with R (correlation coefficient) >0.95 with associated 95% confidence boundaries.

find a degeneration of the structure diversity induced into property diversity whenever either the Carbon or the Nitrogen is used as reference atom. Thus, the free term of the model providing the relationship between the property and the number of distinct factors – see Table 2 – embeds different groups of factors degenerated from the full factorial model when the reference element is changed. On the other hand, a proper choice of the reference atom may reduce the number of factor groups (e.g. 'factors' for the linear association) and therefore may require less number of experiments to catch the whole diversity. For the 'molar volume' (see Supplementary Table), observing only 42 structures (out of 45 different structures) is enough (supposing that the observation has 100% accuracy) to have the coefficients of the equation expressing the molar volume as a function of the type of the atom fully determined.

A model with a correlation coefficient (R) nearby 0.95 (the factors are able to explain ~90% of the variance of the property of interest) could be consider a good explanatory model. Therefore, the models with R > 0.95 were further analyzed. The number of factors in the first model with R > 0.95 varied from 2 to 19 (Table 3), with highest average number of factors observed when carbon was the reference atom (Fig. 4).

An analysis of Table 3 revels that the number of factors is different for different reference atoms with one exception. The exception is observed for energy, where all models had four factors regardless the reference atom. Despite the fact that the average number of factors in the first model with R > 0.95 is higher when carbon is the reference atom, in seven out of ten investigated properties (volume, surface area, ovality, HOMO, LUMO, polarizability, and entropy), the smallest number of factors was observed when carbon is the reference atom.

Different number of factors for the investigated reference atoms could be seen as the diversity induced by the reference atom. The diversity is bad in this case (45 distinct congeners) since relevant associations could be omitted. Furthermore, a property (such as 'volume') has different number of factors according with the reference atom (9/4/2, see Table 3) while these 'imperfect' relationships explained only 'most of' the change in the property of interest (for Vol: 95.48% for B, 98.09% for N, and 95.93% for C). Furthermore, these 'imperfect' relationships led to the imperfect translations of the diversity from the sample into the model.

The smallest number of factors in the model not necessary produce a performing explanatory model, which can be assessed by plotting the calculated vs. estimated property [30,31] (see Fig. 5). For example, the model for LUMO is with the smallest number of factors when C is the reference atom but the graphical representation shows that the best model is the one when N is the reference atom, which is not surprising since this model incorporate a higher number of factors (see Table 3). This behavior is also observed when the number of factors is not necessary the highest, such as is the case of volume (4 factors for N as reference atom, 9 factors for B as reference atom), surface area (5 factors for N as reference), and polarizability (4 factors for N as reference atom). However, this approach is a suitable solution for enthalpy, for example, when a tendency to form clusters is observed regardless the reference atom (see Fig. 5).

The same performances (identical goodness-of-fit) were observed on models for energy regardless of the reference atom, but these models incorporate different factors and interactions:

- Energy_B = −494.86-69.29·(c4 + c1)-149.18·(n4 + n1)-69.06·(c3 + c2)-149.18·(n3 + n2)
- Energy_N = −1091.58 + 149.18·(b4 + b1) + 79.89·(c4 + c1) + 149.18·(b3 + b2) + 80.12·(c3 + c2)
- Energy_C = −771.55-79.89·(n4 + n1) + 69.29·(b4 + b1)-80.12·(n3 + n2) + 69.06·(b3 + b2)

Our study showed that when a relationship is designed to relate properties with structural features for paired data in a population, the information parsimony usually leads to the simpler models [33], models that had incomplete explanatory power [34]. Thus, in these cases, it is always a compromise between increasing the sample size and decreasing the model complexity. Validation of linear models is a critical step in identification of models with prediction abilities [35–39] but the issue of model validation was beyond the aim of our study and was not considered in this manuscript. The study presented in this manuscript was focused on development of a full-factorial design for investigation of the link between features extracted from structure of compounds and several properties. Further studies are currently being conducted in our laboratory to validate the models obtained on the investigated class of one-cage pentagonal faces nanostructure. Furthermore, new *in silico* approaches such as graph derivative indices [40], graph-based approaches [41], matrices descriptors [42–44], adjacency [45,46], chemical colored (weighted) topo-reactivity [47]), and/or similarity or classification approaches [48,49] could join the full-factorial analysis. However, this combination of approaches will require high computation capacity.

4. Conclusions

The applied full factorial design on one-cage pentagonal faces nanostructures showed that the highest correlation coefficient on a non-full model was obtained for energy as property regardless the reference atom (R = 0.9999).

The highest correlation coefficient was obtained when carbon was the reference atom in 67% of cases but these models are not

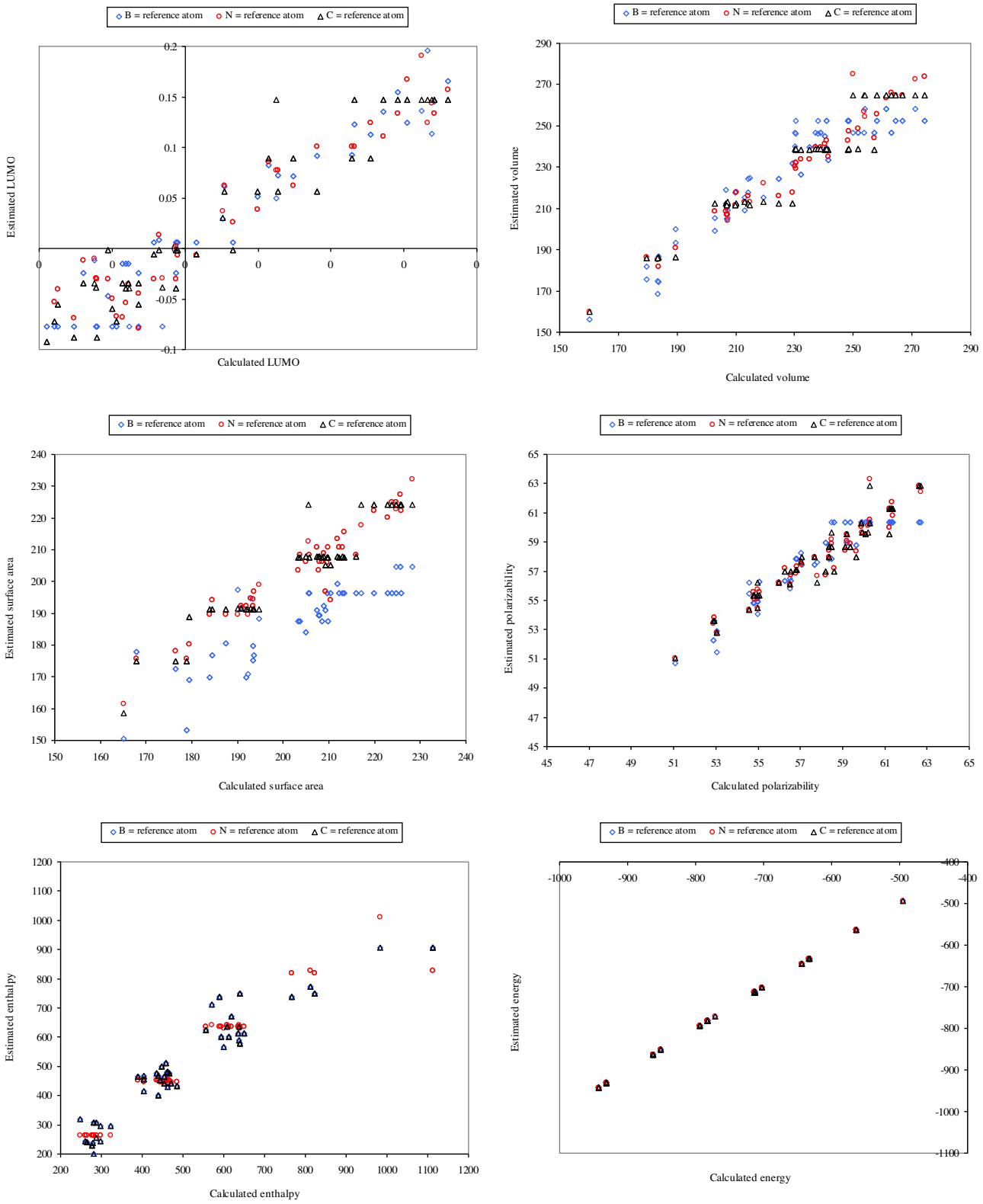


Fig. 5. Estimated versus calculate properties on one-cage pentagonal faces nanostructures.

necessary the best performing models. The complexity of the models increases from boron taken as a reference (simplest models) to carbon taken as reference (convoluted models). Therefore, in order to obtain the simplest expression of the most critical factors, it is also useful to include a factorial analysis of the alternatives for the reference category.

Availability

The implemented executable program is available from the authors upon request.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.md.2016.12.001.

References

[1] R.J. Ternansky, D.W. Balogh, L.A. Paquette, Dodecahedrane, J. Am. Chem. Soc. 104 (1982) 4503–4504.
[2] L.A. Paquette, Dodecahedrane –the chemical transliteration of Plato's universe (a review), Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 4495–4500.
[3] L.A. Paquette, R.J. Ternansky, D.W. Balogh, et al., Total synthesis of dodecahedrane, J. Am. Chem. Soc. 105 (1983) 5446–5450.
[4] J.M. Schulman, T. Venanzi, R.L. Disch, A theoretical study of the dodecahedrane molecule, J. Am. Chem. Soc. 97 (1975) 5335–5339.
[5] J.M. Schulman, R.L. Disch, Theoretical studies of dodecahedrane. 2. Dodecahedrane, inclusion compounds and fluorine derivatives, J. Am. Chem. Soc. 100 (1978) 5677–5681.
[6] T. Clark, T.M. Knox, H. Mackle, et al., Enthalpy of formation of perhydroquinacene and a comment on the strain energy of dodecahedrane, Chem. Commun. 16 (1975) 666–667.
[7] R.L. Disch, J.M. Schulman, Theoretical studies of dodecahedrane: 3. Ab initio studies of dodecahedrane and its inclusion compounds, J. Am. Chem. Soc. 103 (1981) 3297–3300.
[8] L.R. MacGillivray, J.L. Atwood, Structural classification and general principles for the design of spherical molecular hosts, Angew. Chem. 38 (1999) 1018–1033.
[9] R.L. Disch, J.M. Schulman, Heat of formation of dodecahedrane, J. Phys. Chem. 100 (1996) 3504–3506.
[10] F.-L. Liu, L. Peng, J.-X. Zhao, et al., Theoretical study of two C50H40 isomers with three dodecahedrane cages sharing two pentagons, Int. J. Quantum Chem. 103 (2005) 167–175.
[11] S.G. Semenov, M.V. Makarova, A quantum chemical study of diethynyl derivatives of dodecahedrane and buckminsterfullerene in vacuum and in tetrahydrofuran, Opt. Spectrosc. 118 (2015) 46–49.
[12] S. Berski, A.J. Gordon, Z. Latajka, Electron localization function study on the chemical bonding in a real space for tetrahedrane, cubane, adamantane, and dodecahedrane and their perfluorinated derivatives and radical anions, J. Phys. Chem. A 118 (2014) 4147–4156.
[13] R.J. Cross, M. Saunders, H. Prinzbach, Putting helium inside dodecahedrane, Org. Lett. 1 (1999) 1479–1481.
[14] H. Prinzbach, A. Weiler, L.P. Andenberger, et al., Gas-phase production and photoelectron spectroscopy of the smallest fullerene, C_{20}, Nature 407 (2000) 60–63.
[15] S. Tanaka, Remembrance of dodecahedrane, Yuki Gosei Kagaku Kyokaishi/J. Synth. Org. Chem. 60 (2002) 1210–1214.
[16] A.G. Nasibulin, P.V. Pikhitsa, H. Jiang, et al., A novel hybrid carbon material, Nat. Nanotechnol. 2 (2007) 156–161.
[17] F. Wahl, A. Weiler, P. Landenberger, et al., Towards perfunctionalized dodecahedranes-en route to C_{20} fullerene, Chem. Eur. J. 12 (2006) 6255–6267.
[18] G. Banfalvi, Dodecahedrane minibead polymers, RSC Adv. 4 (2014) 3003–3008.
[19] D. Moran, F. Stahl, E.D. Jemmis, et al., Structures, stabilities, and ionization potentials of dodecahedrane endohedral complexes, J. Phys. Chem. A 106 (2002) 5144–5154.
[20] O.A. Gapurenko, T.N. Gribanova, R.M. Minyaev, et al., Hypercoordinate atoms of second-row elements in dodecahedrane endohedral complexes, Russ. Chem. Bull. 56 (2007) 856–862.
[21] Z. Chen, H. Jiao, D. Moran, et al., Aromatic stabilization in heterofullerenes $C_{48}X_{12}$ (X = N, P, B, Si), J. Phys. Org. Chem. 16 (2003) 726–730.
[22] L. Jäntschi, General Chemistry Course AcademicDirect Internet (2013), cited 2015 Sep 20, Available from: http://ph.academicdirect.org/General_Chemistry_Course_v5.pdf.
[23] D.R. Hartree, The wave mechanics of an atom with a non-coulomb central field: part I. Theory and methods, Proc. Cambridge Phil. Soc. 24 (1928) 89–110.
[24] D.R. Hartree, The wave mechanics of an atom with a non-Coulomb central field Part II. Some results and discussion, Proc. Cambridge Phil. Soc. 24 (1928) 111–132.
[25] V. Fock, Naherungsmethode zur Losung des quantenmechanischen Mehrkorper- problems, Z. Phys. 61 (1930) 126–148.
[26] M.S. Møller Chr Plesset, Note on an approximation treatment form many-electron systems, Phys. Rev. 46 (1934) 618–622.
[27] R. Ditchfield, W.J. Hehre, J.A. Pople, Self-consistent molecular-orbital methods: IX. an extended Gaussian-type basis for molecular-orbital studies of organic molecules, J. Chem. Phys. 54 (1971) 724–728.
[28] Jäntschi L. Structure vs. Property: Algorithms and Models. Habilitation thesis (Chemistry, Mathematics, Informatics), Defended at Babeş-Bolyai University on Sept. 24, 2012. cited 2015 Jul 20 Available from: http://www.cnatdcu.ro/wp-content/uploads/2012/06/g_3_Teza_Spam_Summary_Adobe4.pdf.
[29] S.D. Bolboacă, L. Jäntschi, The effect of leverage and/or influential on structure-activity relationships, Comb. Chem. High Throughput Screen. 16 (2013) 288–297.
[30] S.D. Bolboacă, L. Jäntschi, Quantitative structure-activity relationships: linear regression modelling and validation strategies by example, BIOMATH 2 (2013) 1309089.
[31] S.D. Bolboacă, L. Jäntschi, Modelling the property of compounds from structure: statistical methods for models validation, Environ. Chem. Lett. 6 (2008) 175–181.
[32] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, Detecting bad regression models: multicriteria fitness functions in regression analysis, Anal. Chim. Acta 515 (2004) 199–208.
[33] K.T. Kelly, Philosophy of statistics, in: P.S. Bandyopadhyay, M.R. Forster (Eds.), Handbook of the Philosophy of Science, vol. 7, Elsevier, 2011, pp. 983–1024.
[34] K. Kar, S. Arias-Estrada, How to judge predictive quality of classification and regression based QSAR models? in: Z. Ul-Haq, J.D. Madura (Eds.), Frontiers in Computational Chemistry, Volume 2: Computer Applications for Drug Design and Biomolecular Systems, Bentham Science Publishers Ltd., 2015, pp. 71–120.
[35] S.D. Bolboacă, L. Jäntschi, Sensitivity, specificity, and accuracy of predictive models on phenols toxicity J. Comput. Sci. 5 (2014) 345–350.
[36] P. Gramatica, On the development and validation of QSAR models, Methods Mol. Biol. 930 (2013) 499–526.
[37] S.D. Bolboacă, L. Jäntschi, Nano-quantitative structure-property relationship modeling on C_{42} fullerene isomers, J. Chem. 2016 (2016), Article ID 1791756 (8 pages).
[38] L. Jäntschi, S.D. Bolboacă, Szeged matrix property indices as descriptors to characterize fullerenes, Ovidius Univ. Ann. Chem. 27 (2) (2016) 73–80.
[39] M. Tichý, M. Rucki, Validation of QSAR models for legislative purposes, Interdiscip. Toxicol. 2 (2009) 184–186.
[40] O. Martínez-Santiago, Y. Marrero-Ponce, S.J. Barigye, H.L.T. Thu, F. Javier Torres, C.H. Zambrano, J.L. Muñiz Olite, M. Cruz-Monteagudo, R. Vivas-Reyes, L.V. Infante, L.M. Artiles Martínez, Physico-Chemical and structural interpretation of discrete derivative indices on N-tuples atoms, Int. J. Mol. Sci. 17 (6) (2016) (Article number 812).
[41] C.R. Munteanu, V. Aguiar-Pulido, A. Freire, M. Martínez-Romero, A.B. Porto-Pazos, J. Pereira, J. Dorado, Graph-based processing of macromolecular information, Curr. Bioinf. 10 (5) (2015) 606–631.
[42] L. Jäntschi, S.D. Bolboacă, Szeged matrix property indices as descriptors to characterize fullerenes, Ovidius Univ. Ann. Chem. 27 (2) (2016) 73–80.
[43] S.D. Bolboacă, L. Jäntschi, Nano-quantitative structure-property relationship modeling on C42 fullerene isomers, J. Chem. 2016 (2016) (Article ID 1791756).
[44] Putz M.V., Tudoran M.A., Ori O., 2015. Topological Organic Chemistry: From Distance Matrix to Timisoara Eccentricity. 19(3):249–273.
[45] S.C. Basak, Chapter 1. Mathematical chemodescriptors and biodescriptors: background and their applications in the prediction of bioactivity/toxicity of chemicals, in: S. Singh (Ed.), Systems Biology Application in Synthetic Biology, 1st edition, Springer, India, 2016, pp. 117–147.
[46] M.A. Tudoran, M.V. Putz, Molecular graph theory: from adjacency information to colored topology by chemical reactivity, Curr. Org. Chem. 19 (4) (2015) 359–386.
[47] M.V. Putz, O. Ori, F. Cataldo, A.M. Putz, Parabolic reactivity coloring molecular topology: application to carcinogenic PAHs, Curr. Org. Chem. 17 (23) (2013) 2816–2830.
[48] J. Zhou, P. Zhong, T. Zhang, A novel method for alignment-free DNA sequence similarity analysis based on the characterization of complex networks, Evol. Bioinf. Online 12 (2016) 229–235.
[49] I. Borozan, S. Watt, V. Ferretti, Integrating alignment-based and alignment-free sequence similarity measures for biological sequence classification, Bioinformatics 31 (9) (2015) 1396–1404.

Supplementary material:

Elemental factorial study on one-cage pentagonal faces nanostructure congeners

Lorentz Jäntschi, Donatella Bálint, Lavinia L. Pruteanu, and Sorana D. Bolboacă

Factor groups of ~100% determination of variance in molar volume (Vol)

Ref	"B"_fact	"N"_fact	"C"_fact	"B"_coef	"N"_coef	"C"_coef
no	constant(intercept)=true	constant(intercept)=true	constant(intercept)=true	2.50e+2	1.60e+2	2.54e+2
1	c4·c3·c2·c1	b4·b3·b2·b1	n4·n3·n2·n1	-2.19e+1	-3.23e+1	4.80e-1
2	n4·c3·c2·c1+c4·c3·c2·n1	c4·b3·b2·b1+b4·b3·b2·c1	b4·n3·n2·n1+n4·n3·n2·b1	-2.46e+1	-6.68e+0	-4.60e+0
3	c4·n3·c2·c1+c4·c3·n2·c1	b4·c3·b2·b1+b4·b3·c2·b1	n4·b3·n2·n1+n4·n3·b2·n1	-2.89e+1	1.12e+1	-2.10e-1
4	n4·n3·c2·c1+c4·c3·n2·n1	c4·c3·b2·b1+b4·b3·c2·c1	b4·b3·n2·n1+n4·n3·b2·b1	-2.84e+1	8.39e+0	3.10e+0
5	n4·c3·n2·c1+c4·n3·c2·n1	c4·b3·c2·b1+b4·c3·b2·c1	b4·n3·b2·n1+n4·b3·n2·b1+b4·n1+n4·b1	-2.90e+1	7.80e+0	2.51e+0
6	c4·n3·n2·c1	b4·c3·c2·b1	n4·b3·b2·n1	-3.32e+1	1.23e+1	-1.52e+1
7	n4·n3·n2·c1+c4·n3·n2·n1	c4·c3·c2·b1+b4·c3·c2·c1	b4·b3·b2·n1+n4·b3·b2·b1	-2.56e+1	5.08e+0	2.63e+0
8	n4·c3·c2·n1	c4·b3·b2·c1	b4·n3·n2·b1	-4.23e+1	-1.43e+1	2.66e+0
9	n4·n3·c2·n1+n4·c3·n2·n1	c4·c3·b2·c1+c4·b3·c2·c1	b4·b3·n2·b1+b4·n3·b2·b1	-4.35e+1	6.90e-1	6.97e+0
10	n4·n3·n2·n1	c4·c3·c2·c1	b4·b3·b2·b1	-3.23e+1	4.80e-1	-2.19e+1
11	c4·c3·c2+c3·c2·c1	b4·b3·b2+b3·b2·b1	n4·n3·n2+n3·n2·n1	2.06e+1	-8.95e+0	-5.00e-2
12	n4·c3·c2+c3·c2·n1	c4·b3·b2+b3·b2·c1	b4·n3·n2+n3·n2·b1+b3·n2+n3·b2	3.61e+1	-6.90e-1	-4.47e+0
13	c4·n3·c2+c3·n2·c1	b4·c3·b2+b3·c2·b1	n4·b3·n2+n3·b2·n1	2.60e+1	-1.15e+1	-1.24e+0
14	n4·n3·c2+c3·n2·n1	c4·c3·b2+b3·c2·c1	b4·b3·n2+n3·b2·b1	3.93e+1	-1.51e+0	-8.80e-1
15	c4·c3·n2+n3·c2·c1	b4·b3·c2+c3·b2·b1	n4·n3·b2+b3·n2·n1	2.67e+1	-1.32e+1	-8.70e-1
16	n4·c3·n2+n3·c2·n1	c4·b3·c2+c3·b2·c1	b4·n3·b2+b3·n2·b1	4.09e+1	-1.88e+0	-1.55e+0
17	c4·n3·n2+n3·n2·c1	b4·c3·c2+c3·c2·b1	n4·b3·b2+b3·b2·n1	3.39e+1	-9.50e+0	1.29e+1
18	n4·n3·n2+n3·n2·n1	c4·c3·c2+c3·c2·c1	b4·b3·b2+b3·b2·b1	4.13e+1	-4.30e-1	1.36e+0
19	c4·c3·c1+c4·c2·c1	b4·b3·b1+b4·b2·b1	n4·n3·n1+n4·n2·n1	2.32e+1	-9.04e+0	-1.00e-1
20	n4·c3·c1+c4·c2·n1	c4·b3·b1+b4·b2·c1	b4·n3·n1+n4·n2·b1	2.32e+1	-2.30e-1	-2.41e+0
21	c4·n3·c1+c4·n2·c1	b4·c3·b1+b4·c2·b1	n4·b3·n1+n4·b2·n1	2.79e+1	-7.91e+0	3.90e-1
22	n4·n3·c1+c4·n2·n1	c4·c3·b1+b4·c2·c1	b4·b3·n1+n4·b2·b1	2.03e+1	-5.20e-1	4.02e+0
23	c4·c3·n1+n4·c2·c1	b4·b3·c1+c4·b2·b1	n4·n3·b1+b4·n2·n1	2.98e+1	-1.43e+1	4.46e+0
24	n4·c3·n1+n4·c2·n1	c4·b3·c1+c4·b2·c1	b4·n3·b1+n4·b3+b4·n2·b1+b2·n1	4.46e+1	-2.00e-1	-2.28e+0
25	c4·n3·n1+n4·n2·c1	b4·c3·c1+c4·c2·b1	n4·b3·b1+b4·b2·n1	3.44e+1	-7.39e+0	-2.62e+0

26	n4·n3·n1+n4·n2·n1	c4·c3·c1+c4·c2·c1	b4·b3·b1+b4·b2·b1	4.13e+1	-3.80e-1	-1.27e+0
27	c4·c3+c2·c1	b4·b3+b2·b1	n4·n3+n2·n1	-2.28e+1	2.08e+1	1.60e-1
28	n4·c3+c2·n1	c4·b3+b2·c1	b4·n3+n2·b1	-3.34e+1	3.83e+0	-1.60e-1
29	c4·n3+n2·c1	b4·c3+c2·b1	b4·b3+b2·b1	-2.57e+1	1.21e+1	-9.30e-1
30	n4·n3+n2·n1	c4·c3+c2·c1	n4·n2+n3·n1	-2.95e+1	4.90e-1	-1.70e-1
31	c4·c2+c3·c1	b4·b2+b3·b1	b4·n2+n3·b1	-2.06e+1	1.37e+0	2.96e+0
32	n4·c2+c3·n1	c4·b2+b3·c1	n4·b2+b3·n1	-4.08e+1	1.50e-1	7.00e-1
33	c4·n2+n3·c1	b4·c2+c3·b1	b4·b2+b3·b1	-2.74e+1	1.81e+0	1.23e+0
34	n4·n2+n3·n1	c4·c2+c3·c1	n4·n1	-4.89e+1	1.60e-1	3.00e-2
35	c4·c1	b4·b1	b4·b1	-2.22e+1	7.70e-1	2.28e+0
36	n4·c1+c4·n1	c4·b1+b4·c1+c3·c2	n4+n1	-2.87e+1	3.50e-1	-2.36e+1
37	n4·n1	c4·c1	b4+b1	-4.96e+1	3.10e-1	7.50e+0
38	c4+c1	b4+b1	n3·n2	1.30e+1	2.91e+1	-3.00e-2
39	n4+n1	c4+c1	b3·b2	7.31e+0	2.33e+1	-8.41e+0
40	c3·c2	b3·b2	n3+n2	-2.76e+1	1.61e+1	-2.33e+1
41	n3·c2+c3·n2	c3·b2+b3·c2	b3+b2	-2.77e+1	7.14e+0	4.12e+0
42	n3·n2	b3+b2		-3.41e+1	1.92e+1	
43	c3+c2	c3+c2		2.45e+1	2.31e+1	
44	n3+n2			1.74e+0		

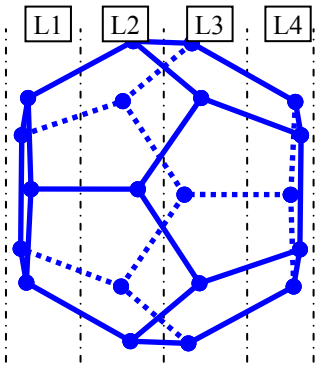
count("fact") = 81 count("fact") = 81 count("fact") = 81

The explanatory equations for molar volume are:

÷ $\mathrm{Vol} \sim \sum_{i=1}^{44} \mathrm{B\_coef}_i \cdot \mathrm{B\_fact}_i$ (for Boron as reference atom)

÷ $\mathrm{Vol} \sim \sum_{i=1}^{43} \mathrm{N\_coef}_i \cdot \mathrm{N\_fact}_i$ (for Nitrogen as reference atom)

÷ $\mathrm{Vol} \sim \sum_{i=1}^{41} \mathrm{C\_coef}_i \cdot \mathrm{C\_fact}_i$ (for Carbon as reference atom)



L1, L2, L3 and L_4 filled with Boron, Carbon or Nitrogen atoms