Computer generation and theoretical study of alicyclic caged structures without small-size cycles

A. V. Evtushenko, M. S. Molchanova, B. B. Smirnov, and V. A. Shlyapochnikov*

N. D. Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences,
47 Leninsky prosp., 117913 Moscow, Russian Federation. Fax: 007 (095) 135 5328

Computer design was conducted for a series of caged alicyclic structures without small cycles. With the help of molecular-mechanical calculations, the enthalpies of formation and the strain energies were determined for the corresponding compounds. The possibility of the existence of highly strained frameworks without small cycles was analyzed.

Key words: caged structures, molecular mechanics, enthalpy of formation, strain energy.

Alicyclic structures have been drawing significant attention from specialists both in experiment and in calculation methods. We mention some of the recent theoretical works: computer-aided construction of promising isomers sharing the molecular formulas of adamantane or wurtzitane,[1—2] as well as their derivatives;[3] experimental and theoretical investigation of nitro derivatives for cubane and some similar structures;[4] geometrical studies of various frameworks.[5—8] However, numerous chemical and geometrical features of caged structures require further analysis. For example, the relationship between the numbers of theoretically possible, chemically probable, and actually synthesized various frameworks is still unknown, although such a study could open new prospects for synthetic chemistry. It is still unclear which factors (naturally, in addition to the size of cycles) affect the strain within the framework; therefore, the problem of determining the structural features and the very existence of highly strained caged frameworks without small-size cycles remains unsolved so far. As to the relationship between the properties of cage-type compounds and the symmetry of their molecules, the information available is still incomplete as well. The existence of such a relationship has been repeatedly stated in the literature (for example, it is believed that compounds with symmetrical molecules form molecular crystals of higher density[1,9]), but a wide field for systematic research is still available.

The goal of this work is to conduct a computer generation (systematic construction) and theoretical analysis of "rigid" alicyclic caged structures having the compositions C_iH_j (where $8 \le i \le 20$ and $12 \le j \le 32$) and containing no small-size (*i.e.*, three- or four-membered) cycles. Since the number of such structures is very great, we confined ourselves to the generation and analysis of only those molecules which satisfy a number of additional requirements: a) all bonds between carbon atoms are single; molecules contain neither side radicals (or acyclic bridges) nor, according to the requirement of rigidity, methylene chains $(-CH_2-)_K$, where $K \ge 2$; b) the target frameworks contain no quaternary carbon atoms; c) the number of methine carbons ($-CH<$) in each structure does not exceed 8.

Results and Discussion

Computer construction of cage hydrocarbons. To construct the isomers, we used the SMOG software.[10] This isomer-generating program, based on a modification of the graph-theoretical algorithm designed by Faradjev[11] and Molodtsov,[12] generates (in a reasonably short time) all real and hypothetical molecules that conform to the given molecular formula and any set of user-specified structural constraints (in our case, this set is specified by the conditions listed above).

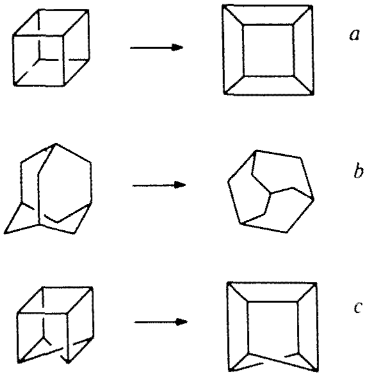


Fig. 1. Examples of (*a*, *b*) planar and (*c*) nonplanar structures. The representation in a plane with a minimum number of intersecting bonds is shown on the right of each structure.

Translated from *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 10, pp. 2430—2437, October, 1996.

Also using the SMOG program, we conducted primary analysis and classification of the resultant molecular structures from the standpoint of their topology.

Firstly, we considered the *symmetry* of caged structures. We can have no definite idea as to the actual symmetry of a molecule (the planes, axes, or center of symmetry) before we calculate its geometry; however, we may characterize the so-called "topological symmetry" of its structural formula. We will describe the topological symmetry by the symmetry number,* also known in graph theory as the order of the automorphism group.[13] If this number is equal to one, the molecule has no elements of symmetry; at the same time, it reaches significant values for highly symmetrical structures (for example, 24 for adamantane, 48 for cubane, or 120 for dodecahedrane). Although molecules with a high symmetry number sometimes have lower actual symmetry (their spatial structure may be significantly distorted compared to the ideal one), this number may be used as a primary estimate of molecular symmetry.[1–3]

The second criterion used for the classification of the molecules generated was their *planarity*. According to the graph-theoretical definition, a molecular structure is planar if it may be drawn in a plane so that no representations of two chemical bonds would intersect[14] (while drawing, one may change the bond lengths and valence angles to any degree, but chemical bonds should not be "broken," and atoms should not be "superimposed"). This definition is illustrated by Fig. 1: the structures of (*a*) cubane and (*b*) adamantane are planar, and the corresponding representations without intersecting bonds are given in the right part of the figure. As the same time, the structure of the so-called Moebius cubane (Fig. 1,*c*) is somewhat similar to that of cubane but nonplanar: each of its representations in a plane contains either superimposed atoms or intersecting chemical bonds.

Evidently, a structure may be represented in the form of a convex polyhedron only if it is planar.

Until recently, the attention of specialists was mainly drawn to planar frameworks. Indeed, all well-known natural and synthetic polyhedral compounds (tetrahedrane, prismanes, cubane, adamantane, wurtzitane, fullerenes, *etc.*) have planar molecular structures. Moreover, the probability of successful synthesis seems to be very low for the overwhelming majority of hypothetical nonplanar structures because of their unfavorable geometry and energy parameters. It may even seem that the existence of nonplanar covalent molecules is forbidden. However, of course, this impression is far from true. For macromolecules, nonplanar forms have been known for a fairly long time — for example, "knotted" molecules of nucleic acids,[15] — and a crown ether with the molecule in the form of a Moebius strip was synthesized for the first time in 1982.[16] As was shown by theoretical analysis,[17] the existence of some nonplanar caged molecules having very complex topology may be quite probable from the chemical standpoint, and, moreover, these frameworks may be even unstrained. Among theoretical studies, also notable is the review,[18] in which dozens of promising hypothetical frameworks are considered, some of them being nonplanar. Therefore, we decided that nonplanar structures generated by SMOG should be analyzed alongside with planar ones rather than excluded from consideration.

The geometry and thermochemistry of the molecules were studied by the PC4 program, which is based on the method of molecular mechanics (MM).[19] Certainly, the accuracy ensured by this technique is low compared to that of quantum chemical methods: MM often allows one to obtain only semi-qualitative estimates for enthalpies of formation (in the ideal gas state) of different compounds. However, such an accuracy is sufficient for this study, because our aim at the first stage of investigations was to reveal only qualitative (rather than quantitative) relationships between the structures and properties of frameworks. Hence, we wanted to analyze the greatest possible number of various frameworks rather than perform high-accuracy calculations for a few of the structures generated. The MM method is suitable for such a wide-scale analysis because of its high performance.

In addition, almost all molecular-mechanical programs currently used are based on the modification of the MM method that yields satisfactory results for polyhedral molecules.[19,20] For example, the experimental and calculated enthalpies of formation are equal for adamantane (-0.23 kcal g^{-1}); for diamantane, these values amount to -0.19 and -0.18 kcal g^{-1}, respectively; *etc.* *

Finally, the MM method directly provides information on such an important characteristic of molecules as their strain energy.

Calculations of molecular energy and topology. Before computer generation of caged molecules, we may make a number of conclusions as to the possible compositions of target structures. This procedure will enable us to choose the molecular formulas to be used for isomer generation.

As directly follows from conditions (a) and (b) of the previous section, the target molecules consist only of

---

*We may define this value as the number of permutations between atoms in a molecule that preserve its connectivity — *i.e.*, the number of permutations that "transform" the molecule into an identical structure. Usually only permutations of skeleton atoms are considered, and hydrogens that are bound to them are disregarded. For example, the symmetry number equals 2 in the case of ethane (the atoms may be either "left at their positions" or "interchanged"), 4! = 24 in the case of 2,2-dimethylpentane (any permutations of the four primary carbons preserve the connectivity of the molecule), 8 in the case of cyclobutane (atoms 1 and 3, 2 and 4 may interchange their positions, as well as the pairs 12 and 34, *etc.*). For molecules without any topological symmetry (for example, 2-methylnorbornane), the symmetry number is equal to 1.

*Unfortunately, the experimental data on the heats of formation of caged compounds (for the ideal gas state) are very scarce and often inconsistent.

methylene ($-CH_2-$) and methine ($>CH-$) groups. The number of C—C bonds in a molecule that consists only of the $-CH_2-$ and $-CH<$ fragments is equal to

$$P = N_{CH_2} + 3/2\ N_{CH}, \qquad (1)$$

where N_{CH_2} and N_{CH} are the numbers of methylene and methine groups, respectively. Hence, the number of CH-fragments is always even.

Secondly, the absence of CH_2-CH_2 chains implies that each CH_2-group is bound to two CH-fragments. At the same time, if the number of methine groups N_{CH} is given, the maximum possible N_{CH_2} value is reached when each methine group is bound to three methylene fragments (for example, as in adamantane). Accordingly,

$$N_{CH_2} \le 3/2 N_{CH}. \qquad (2)$$

If we consider only the compositions for which planar structures are possible in principle (as was mentioned above, only a planar molecule may be represented by a convex polyhedron), then the maximum possible number of CH_2-groups for the given N_{CH} may be found according to Euler's formula[14]

$$F + V - E = 2, \qquad (3)$$

where F is the number of faces in the polyhedron that represents the caged structure (each face represents a cycle), V is the number of vertices (carbon atoms), and E is the number of edges (C—C bonds). In our case, $V = N_{CH_2} + N_{CH}$, and E is determined by formula (1). Then,

$$F = 2 + N_{CH} / 2. \qquad (4)$$

Since each bond is shared by two faces (cycles), it is easy to find that the "average" size of a cycle is equal to

$$C = (3N_{CH} + 2N_{CH_2})/F. \qquad (5)$$

If this value is smaller than 5, then, evidently, a convex polyhedron without small cycles cannot be constructed for the given composition.

Thus, planar structures that satisfy all the aforementioned constraints may exist only if the inequality (2) and the condition are satisfied.

$$N_{CH_2} \ge 5 - N_{CH} / 4 \qquad (6)$$

Certainly, there still remains the possibility to construct nonplanar structures at smaller N_{CH_2} values. We should note at once that only two such structures were found by the program (the frameworks **4** and **32** in Fig. 2), and those were highly strained. All other compositions considered in the generation satisfied the inequalities (2) and (6), as well as condition (c) of the previous section.

The results of our calculations for 21 molecular formulas (from C_8H_{12} to $C_{20}H_{32}$) are shown in Table 1. Apparently, the number of isomers generated by SMOG is comparatively small: less than one thousand for all the

Table 1. Molecular formulas of isomer families, numbers of CH and CH_2 groups for each of them (N_{CH} and N_{CH_2}), and numbers of isomers generated (K) classified by their symmetry numbers and planarity (K_p and K_n are the numbers of planar and nonplanar isomers, respectively).

Formula	N_{CH}	N_{CH_2}	Isomer distribution by symmetry numbers.										K	K_p	K_n
			1	2	4	6	8	12	16	24	48	72			
C_8H_{12}	4	4					1						1	1	0
C_9H_{14}	4	5			1								1	1	0
$C_{10}H_{16}$	4	6								1			1	1	0
C_9H_{12}	6	3						1					1	0	1
$C_{10}H_{14}$	6	4		3			1						4	1	3
$C_{11}H_{16}$	6	5	1	6			1						8	4	4
$C_{12}H_{18}$	6	6	2	5	1			3					11	7	4
$C_{13}H_{20}$	6	7	1	2	4								7	5	2
$C_{14}H_{22}$	6	8		1	1		1						3	2	1
$C_{15}H_{24}$	6	9						1				1	2	1	1
$C_{10}H_{12}$	8	2					1						1		1
$C_{11}H_{14}$	8	3	1	3	1	1							6	1	5
$C_{12}H_{16}$	8	4	14	14	2		1		1				32	9	23
$C_{13}H_{18}$	8	5	53	34	3								90	27	63
$C_{14}H_{20}$	8	6	114	38	15	1		2					170	67	103
$C_{15}H_{22}$	8	7	139	64	10								213	100	113
$C_{16}H_{24}$	8	8	115	57	13		2		3				190	106	84
$C_{17}H_{26}$	8	9	57	45	9	3	2	1					117	72	45
$C_{18}H_{28}$	8	10	15	22	11		2		2				52	35	17
$C_{19}H_{30}$	8	11		1	1		1						3	2	1
$C_{20}H_{32}$	8	12			1			1	2		1		5	3	2
Total													918	445	473

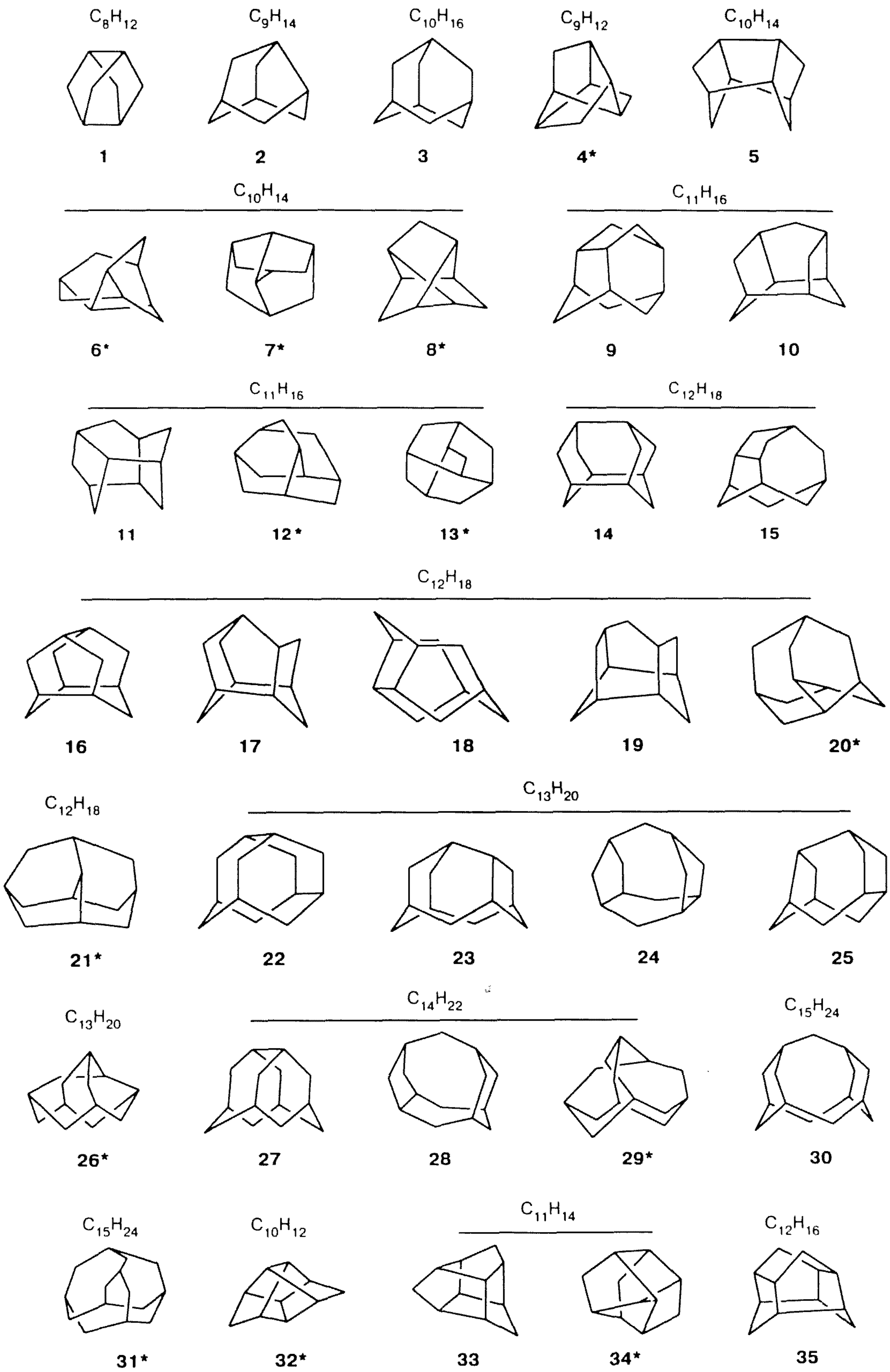
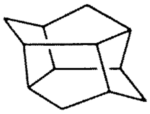


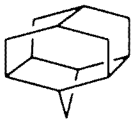
Fig. 2. (to be continued).

$C_{12}H_{16}$

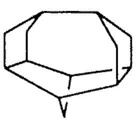


36

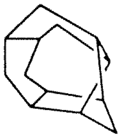
$C_{13}H_{18}$



37

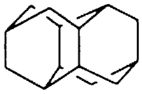


38



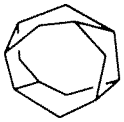
39*

$C_{14}H_{20}$

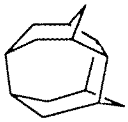


40

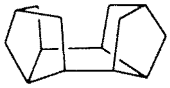
$C_{14}H_{20}$



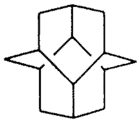
41



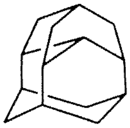
42



43

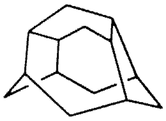


44



45*

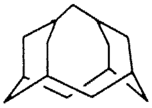
$C_{14}H_{20}$



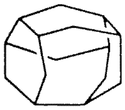
46*

$C_{15}H_{22}$

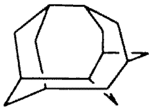
47



48

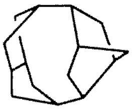


49

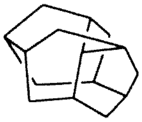


50

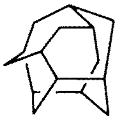
$C_{15}H_{22}$



51

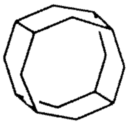


52*

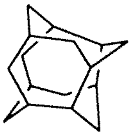


53*

$C_{16}H_{24}$

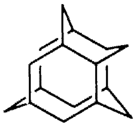


54

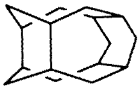


55

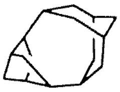
$C_{16}H_{24}$



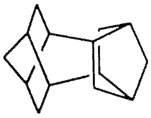
56



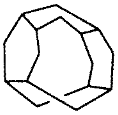
57



58

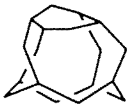


59



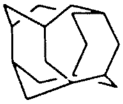
60*

$C_{16}H_{24}$

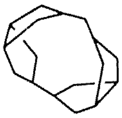


61*

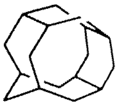
$C_{17}H_{26}$



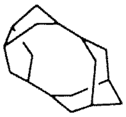
62



63

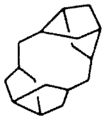


64

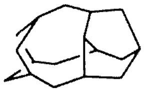


65

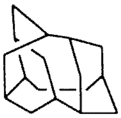
$C_{17}H_{26}$



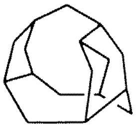
66



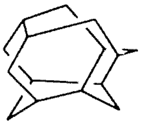
67*



68*



69*



70*

Fig. 2. (continued).

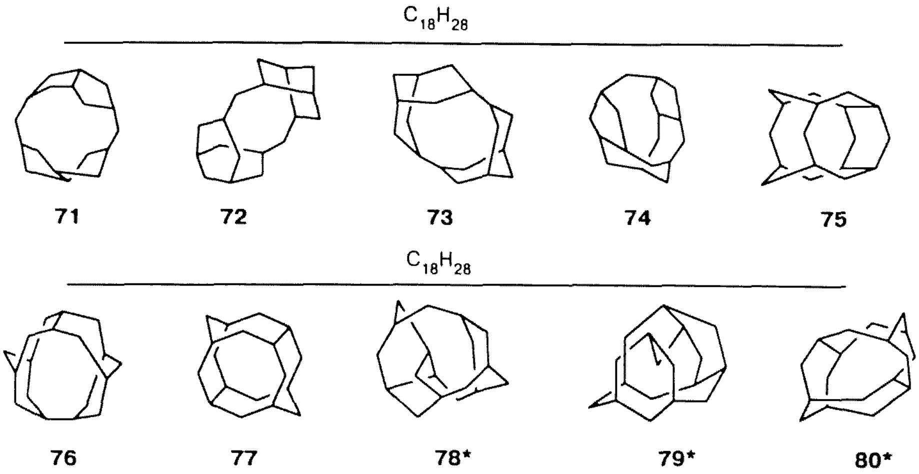


Fig. 2. Some of the planar and nonplanar hypothesized cage-type molecules generated by the SMOG software. The enthalpies of formation and strain energies are given in kcal g^{-1}. Nonplanar structures are marked by asterisks.

Compond	$\Delta_f H°$	E_{st}	Compond	$\Delta_f H°$	E_{st}	Compound	$\Delta_f H°$	E_{st}	Compound	$\Delta_f H°$	E_{st}	Compound	$\Delta_f H°$	E_{st}
1	0.14	0.41	**17**	−0.05	0.20	**33**	0.08	0.30	**49**	−0.02	0.25	**65**	0.02	0.32
2	−0.12	0.18	**18**	−0.02	0.22	**34**	2.97	3.20	**50**	0.03	0.31	**66**	0.03	0.33
3	−0.23	0.07	**19**	0.08	0.33	**35**	0.03	0.27	**51**	0.09	0.37	**67**	0.11	0.40
4	2.69	2.93	**20**	1.68	1.96	**36**	0.002	0.24	**52**	1.00	1.28	**68**	1.02	1.32
5	0.06	0.32	**21**	1.97	2.26	**37**	−0.07	0.19	**53**	2.02	2.30	**69**	1.08	1.37
6	1.20	1.45	**22**	−0.12	0.18	**38**	0.01	0.26	**54**	−0.05	0.24	**70**	1.39	1.69
7	1.27	1.52	**23**	−0.10	0.19	**39**	1.80	2.06	**55**	−0.04	0.25	**71**	−0.07	0.24
8	1.99	2.24	**24**	−0.09	0.20	**40**	−0.18	0.09	**56**	−0.02	0.27	**72**	−0.03	0.27
9	−0.06	0.22	**25**	−0.05	0.24	**41**	−0.07	0.19	**57**	−0.01	0.28	**73**	−0.03	0.27
10	−0.02	0.25	**26**	1.41	1.71	**42**	−0.04	0.23	**58**	0.06	0.35	**74**	−0.007	0.29
11	0.09	0.36	**27**	−0.10	0.20	**43**	0.005	0.27	**59**	0.18	0.41	**75**	0.03	0.34
12	1.30	1.57	**28**	−0.08	0.22	**44**	0.24	0.51	**60**	0.98	1.27	**76**	0.04	0.35
13	2.12	2.39	**29**	1.62	1.93	**45**	0.98	1.25	**61**	1.10	1.44	**77**	0.05	0.35
14	−0.12	0.17	**30**	−0.07	0.23	**46**	1.74	2.01	**62**	−0.04	0.25	**78**	0.49	0.79
15	−0.11	0.17	**31**	2.09	2.40	**47**	−0.03	0.25	**63**	−0.04	0.25	**79**	0.92	1.20
16	−0.08	0.21	**32**	2.44	2.66	**48**	−0.02	0.26	**64**	0.004	0.30	**80**	1.20	1.50

formulas, and the fraction of planar structures is somewhat smaller than 50 percent (445). In particular, this result means that the structural class of caged hydrocarbons is rather narrow, especially if the above structural constraints are considered: for example, the total number of theoretically possible structural isomers for the single molecular formula $C_{10}H_{16}$ amounts to 24938 (*e.g.*, see the paper[21]), but, among these structures, adamantane is the only one satisfying the conditions formulated in this paper (see Table 1).

Figure 2 shows some of the planar and nonplanar frameworks generated (the numbers at the latter structures are marked by asterisks). One may note that the molecules in this figure include both well-known structures — such as bis-noradamantane (**1**), noradamantane (**2**), adamantane (**3**), wurtzitane (**14**), isowurtzitane (**16**), diamantane, or congressane (**40**), *etc.* — and currently unknown hypothetical frameworks. For all structures with $N_{CH} < 8$ and highly symmetrical (see above) structures with $N_{CH} = 8$, we conducted geometry optimization and calculated their enthalpies of formation for the ideal gas state ($\Delta_f H°$) and strain energies (E_{st}) by the MM method.

According to our computed results, the whole set of structures for which the MM calculations were conducted may be formally divided into two groups: the first one includes unstrained and slightly strained frameworks; the other, strained and highly strained ones. Since the enthalpy of formation is determined by the total energy of bonds and the strain energy of the molecule, we may assume that the first group and the second one comprise structures with relatively low (smaller than 0.3 kcal g^{-1}) and high (0.3 kcal g^{-1} and above) enthalpies of formation, respectively.

It turned out that the above classification of structures by their energy and strain almost completely coincides with their formal classification by the property of planarity. Low-strain structures are planar frameworks without small cycles, whereas strained and highly strained structures are nonplanar.

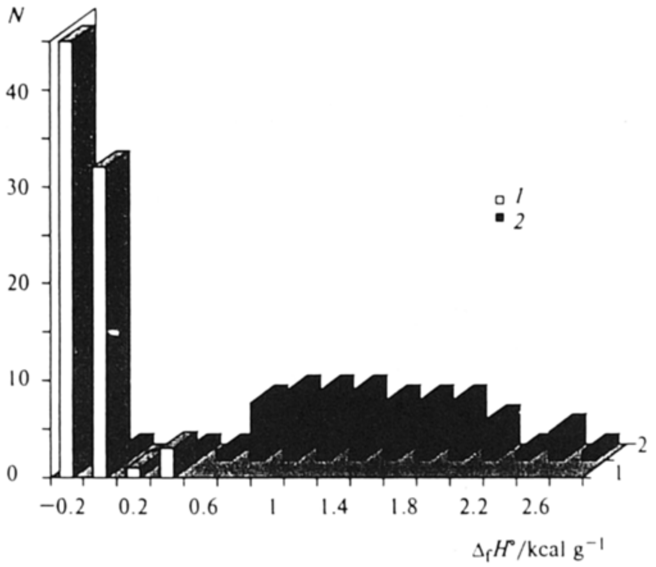


Fig. 3. The distribution of planar (*1*) and nonplanar (*2*) structures by their calculated specific enthalpy of formation.

The bar chart (Fig. 3) shows the distribution of the planar and nonplanar frameworks analyzed in this study by their specific enthalpies of formation. It is notable that almost all planar frameworks are concentrated within the sufficiently narrow energy range from −0.2 to 0.2 kcal g^{-1}, whereas the distribution of nonplanar structures by their calculated enthalpy is much wider and "shifted" toward higher values: from 0.4 to 2.8 kcal g^{-1}. There are also some outliers: for example, the calculated specific enthalpy of the hypothesized nonplanar structure **67** in Fig. 2 is unexpectedly low (0.11 kcal g^{-1}). However, it is possible that this result is due to insufficient accuracy of our calculations.

So, we may make a number of preliminary conclusions. Theoretically, alicyclic caged structures without multiple bonds and small-size cycles may be highly strained if they are nonplanar. The problem of the corresponding substances remains open and should be solved separately for each of the hypothesized compounds. If such syntheses are successful, we will be able to state that an essentially new class of energy-capacious substances has appeared: nonplanar frameworks without small cycles and their derivatives.

Naturally, the probability of successful synthesis is higher for the less strained molecules among the nonplanar ones. Although the theoretical enthalpies of formation are fairly great for the corresponding hypothetical substances (as a rule, 0.4 kcal g^{-1} or higher), they still remain reasonable. Indeed, the $\Delta_f H°$ values for such well-known compounds as cyclopropane and spiropentane amount to 0.3 and 0.65, respectively, whereas cubane, bicyclobutane, and tetrahedrane are even more strained.

Until the above nonplanar frameworks are synthesized, we do not know the limits of the applicability of the MM method with respect to calculations of strained nonplanar structures. It will be appropriate to conduct quantum-chemical calculations for the molecules in question.

The use of computational methods makes it possible to select the promising structures from a set of hypothetical molecules and to reject (in advance) the compounds for which the probability of successful synthesis appears to be low because of their extremely high strain energy (for example, exceeding that of prismane and cubane).

References

1. T. S. Pivina, M. S. Molchanova, V. V. Shcherbukhin, and N. S. Zefirov, *Propellants, Explosives, Pyrotechnics*, 1994, **19**, 286.
2. T. S. Pivina, E. A. Arnautova, M. S. Molchanova, and V. V. Shcherbukhin, *J. Phys.* III, 1995, **5**, 505.
3. T. S. Pivina, V. V. Shcherbukhin, M. S. Molchanova, and N. S. Zefirov, *Propellants, Explosives, Pyrotechnics*, 1995, **20**, 144.
4. J. Alster, S. Iyer, and O. Sandus, *NATO ASI Ser., Ser. C, 309 (Chem. Phys. Energ. Mater.)*, 1990, 641.
5. B. W. Clare and D. L. Kepert *J. Math. Chem.*, 1991, **6**, 325
6. R. B. King, *J. Math. Chem.*, 1991, **7**, 51.
7. Y.-M. Wang, Ch. Chen, and Sh.-S. Cheng, *Huoyau Jishu*, 1994, **10**, 11.
8. Y. Wang, *The Synthesis and Chemistry of Polycyclic Cage Compounds. Diss. Abstr. Int. B*, 1995, **56** (1), 249.
9. A. T. Nielsen, in *Chemistry of Energetic Materials*, Eds. G. A. Olah and D. R. Squire, Academic Press, 1991.
10. M. S. Molchanova, V. V. Shcherbukhin, and N. S. Zefirov, *J. Chem. Inf. Comput. Sci.*, 1996, **36** (4), 888.
11. I. A. Faradjev, in *Algorithmic Investigations in Combinatorics*, Nauka, Moscow, 1978, 11 (in Russian).
12. S. G. Molodtsov, *Commun. Math. Chem.* (MATCH), 1994, **30**, 213.
13. M. I. Nechepurenko, V. K. Popkov, S. M. Mainagashev, *et al., Algorithms and Programs for Solution of Graph and Network Problems*, Nauka, Moscow, 1990 (in Russian).
14. F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 1969.
15. V. I. Sokolov, *Introduction to Theoretical Stereochemistry*, Nauka, Moscow, 1982 (in Russian).
16. D. M. Walba, R. M. Richards, and R. C. Haltiwanger, *J. Am. Chem. Soc.*, 1982, **104**, 3219.
17. S. Elk, *J. Chem. Inf. Comput. Sci.*, 1990, **30**, 69.
18. N. S. 7efirov, S. S. Tratch, and O. S. Chizhov, *Itogi Nauki i Tekhniki, Ser. Org. Chem.*, **3**, VINITI, Moscow, 1979, 90 pp. (in Russian).
19. T. Clark, *A Handbook of Computational Chemistry*, J. Wiley and Sons, New York, 1985.
20. E. Eugler, J. Andose, and P. Schleyer *J. Am. Chem. Soc.*, 1973, **95**, 8005.
21. R. Grund, A. Kerber, and R. Laue, *Commun. Math. Chem.* (*MATCH*), 1992, **27**, 87.

Received October 9, 1995;
in revised form May 17, 1996