

CORRELATION BETWEEN ENERGIES OF PROPER
FULLERENES AND THEIR TOPOLOGICAL INVARIANTS.
PART I. FULLERENES WITH ABUTTING PENTAGONS

Alexandru T. Balabar,^{a*} Ovidiu Ivanciuc,^a and Darko Babic^b

^a University "Politehnica" of Bucharest, Department of Organic Chemistry,
Oficiul 12 CP 243, 78100 Bucharest, Romania

^b Rudjer Boskovic Institute, Zagreb, Croatia

ABSTRACT. For all constitutional isomers of proper fullerenes with $N = 50$ and 60 carbon atoms, correlations for molecular mechanics-derived energies with r around 0.98 were developed by multilinear regression using four to six molecular descriptors. The descriptors are based on the sizes and disposition of polygons surrounding each edge. By treating separately the 1812 C_{60} isomers and the 271 C_{50} isomers, standard deviations around 4.2 and 3.3 kcal/mol, respectively, were obtained, if one combines all these isomers and if one includes the number of carbon atoms as an additional descriptor, standard deviations s of about 4.0 kcal/mol result. For systems with $N < 50$ carbon atoms, the correlations are less good.

* To whom correspondence should be addressed.

Introduction

Proper fullerenes are defined as consisting of twelve pentagons and any number $(N/2-10)$, except 1) of hexagons. Each of their N carbon atoms ($N \geq 20$) has connectivity three, i.e. the cage is equivalent to a trivalent (cubic) graph. It is necessary to define the term "proper fullerenes" because Herndon,¹ Babic, Trinajstić,^{2,3} Randić,⁴ Scuseria and their coworkers⁵ have discussed fullerenes whose ring sizes may differ from 5 and 6.

Kroto, Smalley, Curl and their coworkers⁶ discovered the fullerenes and ascribed the high stability of fullerenes with $N = 60$ and 70 to the fact that they have isolated (non-abutting) pentagons; the periphery of a pair of abutting pentagons is an eight-membered ring which is antiaromatic. Schmalz, Klein et al.⁷⁻¹⁰ as well as Manolopoulos¹¹ and Fowler¹² found by computer generation that C_{60} has 1812 constitutional isomers, one of which obeys the Isolated Pentagon Rule (IPR); also, C_{70} has only one IPR isomer.¹³⁻¹⁶

In a previous paper,¹⁷ some topological invariants of fullerenes were discussed. The simplest are p , the number of edges shared between pairs of abutting pentagons, ($0 \leq p \leq 20$) and q , the number of vertices shared by pentagon triples ($0 \leq q \leq 10$). Table 1 presents the numbers of constitutional isomers of all proper fullerene isomers with $q = 0$, $N \leq 60$, and Table 2 the numbers of preferable fullerenes with $p = q = 0$ and $60 \leq N \leq 96$. A table showing all the 1812 C_{60} isomers in the terms of p and q may be seen in references^{18,19} (if enantiomerism would be included, there would be 3532 isomers). Chirality will be ignored henceforth throughout our discussion.

The partition of the N vertices of fullerenes according to the sizes of the three polygons in the immediate vicinity of each vertex leads to four types of vertices, $A - D$, as seen in the upper row of Fig. 1.

Table 1. Isomer counts for proper fullerenes with $N=20$ to 60 vertices.

N	Number of isomers	No. of isomers with $q=0$	N	Number of isomers	No. of isomers with $q=0$
20	1	0	42	45	5
22	0	0	44	89	16
24	1	0	46	116	20
26	1	0	48	199	43
28	2	0	50	271	62
30	3	0	52	437	123
32	6	0	54	580	174
34	6	0	56	924	300
36	15	2	58	1205	442
38	17	1	60	1812	704
40	40	5			

Table 2. Isomer counts for the 558 preferable proper fullerenes with $N=60$ to 96 vertices.

N	60	70	72	74	76	78	80	82
Total counts	1	1	1	1	2	5	7	9
N	84	86	88	90	92	94	94	96
Total counts	24	19	35	46	86	134	187	

In the set of descriptors $\{N, p, q, A, B, C, D\}$, the following relationships hold:

$$A = q \quad (1)$$

$$N = A + B + C + D \quad (2)$$

$$B = 2p - 3q, \text{ hence } p = (B + 3A)/2 \quad (3)$$

A table showing all the 1812 C_{60} isomers in terms of A ($0 \leq A \leq 10$) and B ($0 \leq B \leq 24$) may be seen in references.^{17,20} Table 3 presents the energies of the 1812 constitutional isomers of C_{60} in terms of the p, q parameters. One may see that the p, q parameters operate a partition of these isomers into 42 sets, each of these having $n(p, q)$ isomers; three of these sets contain over 200 isomers, and

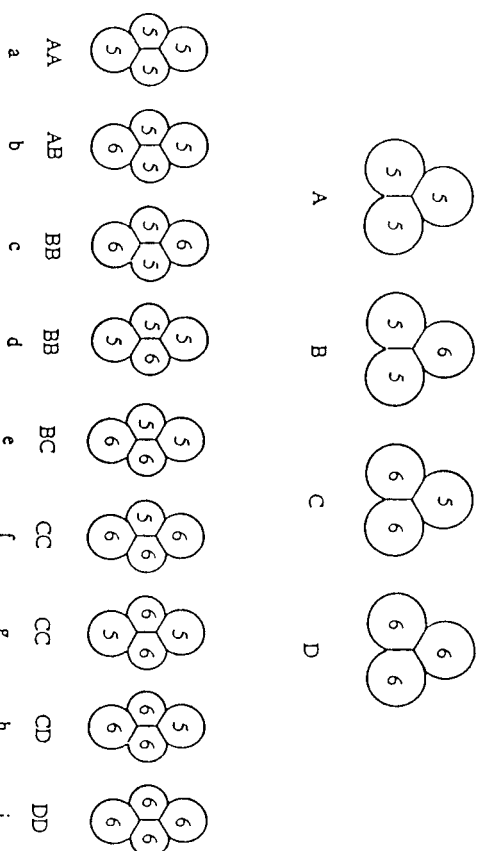


FIG. 1. Upper row: the four types of a vertex in proper fullerenes according to the sizes of the three polygons meeting at this central vertex; lower row: the nine types of an edge in proper fullerenes according to the types of its endpoints and to the sizes of the four polygons in its immediate vicinity.

within a set energies E may differ by $\Delta E = E_{\max} - E_{\min}$ more than 40 kcal/mol; seven sets, indicated in boldface characters, contain just one "unique" isomer, which will be discussed in the next paragraph.

For proper fullerenes possessing abutting pentagons, better discrimination than the above simple invariants may be found. In the previous paper¹⁷ several invariants were explored; some of these were found to afford complete discrimination among all 1812 isomers.

Austin et al.^{21, 22} published recently a thorough study of structural motifs allowing to calculate energies of proper fullerenes. The best fit for the 1812 C_{60} isomers was found for the second moment of the hexagon-neighbor index signature, when the standard deviation of the linear fit was 12.7 kcal/mol.

Table 3. MM Energy partition (in kcal/mol) of the 1812 C_{60} isomers according to p and q .

p	q	$n(p,q)$	E_{ave}	E_{min}	E_{max}	RMS	ΔE
0	0	1	267.63	267.63	267.63	0.000	0.00
2	0	1	282.22	282.22	282.22	0.000	0.00
3	0	3	288.40	287.20	289.20	0.864	2.00
4	0	17	296.08	290.02	300.55	2.900	10.53
5	0	81	304.66	294.85	314.81	4.076	19.96
5	1	5	303.70	298.20	310.24	3.885	12.04
6	0	215	315.49	305.10	333.46	4.702	28.36
6	1	39	311.08	303.21	318.75	4.057	15.54
7	0	210	328.18	318.15	345.67	4.992	27.52
7	1	147	320.82	309.01	335.75	4.985	26.74
7	2	6	318.97	313.09	327.90	4.637	14.81
8	0	145	341.18	328.21	366.13	5.993	37.92
8	1	214	333.36	321.60	349.07	5.444	27.47
8	2	54	328.45	318.62	340.77	5.723	22.15
9	0	23	356.03	340.66	373.96	9.247	33.30
9	1	132	347.08	334.21	368.93	6.985	34.72
9	2	131	339.33	328.27	356.32	6.482	28.05
9	3	11	338.01	327.22	348.05	6.373	20.83
10	0	7	374.97	364.08	397.72	10.906	33.64
10	1	28	360.48	344.16	380.12	8.390	35.96
10	2	116	352.97	337.39	374.06	8.064	36.67
10	3	42	347.65	335.66	359.79	7.245	24.13
10	4	4	350.22	341.77	357.55	5.834	15.78
11	1	1	377.31	377.31	377.31	0.000	0.00
11	2	31	364.49	347.98	379.33	8.158	31.75
11	3	54	359.29	340.79	381.29	9.156	40.50
11	4	10	355.09	339.23	368.29	9.059	29.06
12	0	1	390.72	390.72	390.72	0.000	0.00
12	2	6	386.90	381.31	391.17	3.781	9.86
12	3	16	376.28	359.99	390.71	7.978	30.72
12	4	25	367.41	345.90	385.12	9.291	39.22
12	5	2	369.46	368.89	370.02	0.565	1.13
13	3	2	395.83	393.73	397.93	2.100	4.20
13	4	10	382.63	371.91	396.25	8.373	24.34
13	5	7	376.57	359.09	390.34	9.955	31.25
14	4	3	412.21	406.95	417.72	4.400	10.77
14	5	5	394.57	383.58	401.00	7.020	17.42
14	6	2	398.54	398.07	399.01	0.470	0.94
15	6	2	385.24	372.44	398.05	12.805	25.61
16	6	1	431.22	431.22	431.22	0.000	0.00
18	8	1	454.36	454.36	454.36	0.000	0.00
20	10	1	477.44	477.44	477.44	0.000	0.00

The world of C_{60} isomers - some examples.

In the Atlas of Fullerenes²³ one finds drawings of pentagon-abutting fullerenes with $N = 20$ through 50, and IPR fullerenes with $N = 60$ through 100. Seven isomers chosen among all 1812 C_{60} isomers are described below and displayed in Fig.2 are unique in terms of $\{p, q\}$, or $\{A, B\}$ descriptors. Buckminsterfullerene ($I_h C_{60}$) with $p = q = A = B = 0$ is the lowest-energy isomer, because it obeys the IPR rule. The highest-energy isomer is the "thin-sausage" graphitic cylinder capped by hemidodecahedra; it has $p = 20, q = A = B = 10$. The "thick-sausage" graphitic cylinder (capped by two hexagons surrounded by six pentagons each) has $p = 12, q = A = 0, B = 24$. One Stone-Wales rearrangement²⁴ converts buckminsterfullerene into the isomer with $p = 2, q = A = 0, B = 4$, which still preserves quasi-spherical shape and is next lowest in energy after buckyball. One generalized Stone-Wales rearrangement²⁵ converts the "thick-sausage" isomer into the isomer with $p = 16, q = A = 6, B = 14$ which is elongated and chiral, like the isomer with $p = 18, q = A = 8, B = 12$ which is next in energy after the "thin-sausage" isomer. The last unique isomer with $p = 11, q = A = 1, B = 19$ is also chiral but is less elongated. Another generalization of the Stone-Wales rearrangement may also be mentioned.²⁶

A few other non-unique but symmetrical C_{60} isomers are shown in Fig.3: a trigonal-symmetry isomer with $p = 15, q = A = 6, B = 12$; the "pillow" isomer with two coronene units and six pairs of abutting pentagons has hexagonal symmetry and $p = 6, q = A = 0, B = 12$; a trigonal-symmetry isomer with one coronene face, and $p = 6, q = A = 0, B = 12$; and an isomer with $p = 4, q = A = 0, B = 8$, having four-fold symmetry with four pairs of abutting pentagons and four isolated pentagons.

Partition of fullerene edges into edge types

On considering the sizes of the four polygons in the immediate vicinity of each of the $3N/2$ edges, one obtains a partition of these edges into nine types denoted by letters $a - i$, as seen in the lower row of Fig. 1.

Among the integer numbers mentioned above, the following relationships hold:

$$3N/2 = a + b + c + d + e + f + g + h + i \quad (4)$$

$$3A = 2a + b \quad (5)$$

$$3B = b + 2(c + d) + e \quad (6)$$

$$3C = e + 2(f + g) + h \quad (7)$$

$$3D = h + 2i \quad (8)$$

$$c = p - a - b \quad (9)$$

The two polygons sharing both endpoints of the edge have the first priority in ordering the invariants a through i , relative to the two polygons having only one endpoint of the edge. The two endpoints may each be of type $A - D$, (with some restrictions, as seen in Fig.1) but there are two non-identical situations for the BB and CC type of edge.

Additional topological invariants

Starting from the idea that agglomeration of pentagons leads to pronounced positive curvature, hence to an increase of energy due to the lower solid angle around the sp^2 -hybridized carbon atom, we include $q = A$ as an additional parameter, i.e. the number of vertices of type A.

One the other hand, agglomeration of hexagons, despite the favorable planar disposition of bonds around sp^2 -hybridized carbon atoms, leads to graphitic zones which tend to be planar, but whose cylindrical folding also leads to an

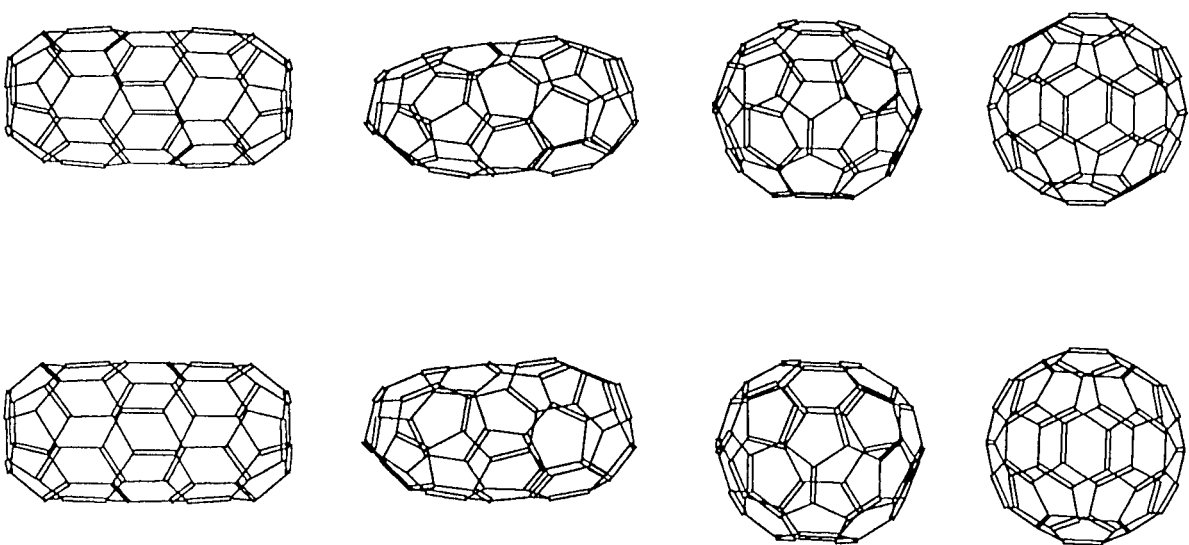


FIG. 2. Stereo-views of the unique C_{60} isomers (from top to bottom): 2A, $p = q = 0$; 2B, $p = 2$, $q = 0$; 2C, $p = 11$, $q = 1$; 2D, $p = 12$, $q = 0$.

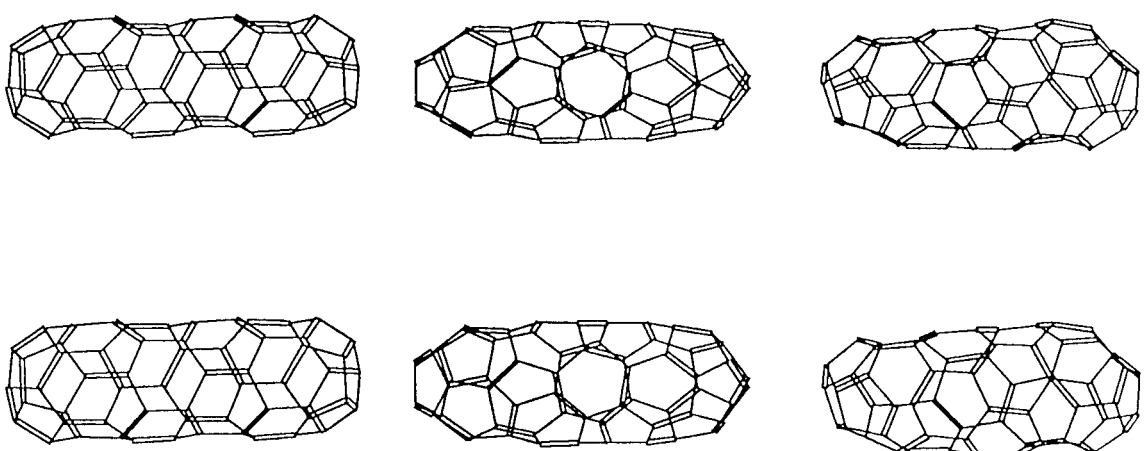


FIG. 2 (continued). Stereo-views of the unique C_{60} isomers (from top to bottom): 2E, $p = 16$, $q = 6$; 2F, $p = 18$, $q = 8$; 2G, $p = 20$, $q = 10$.

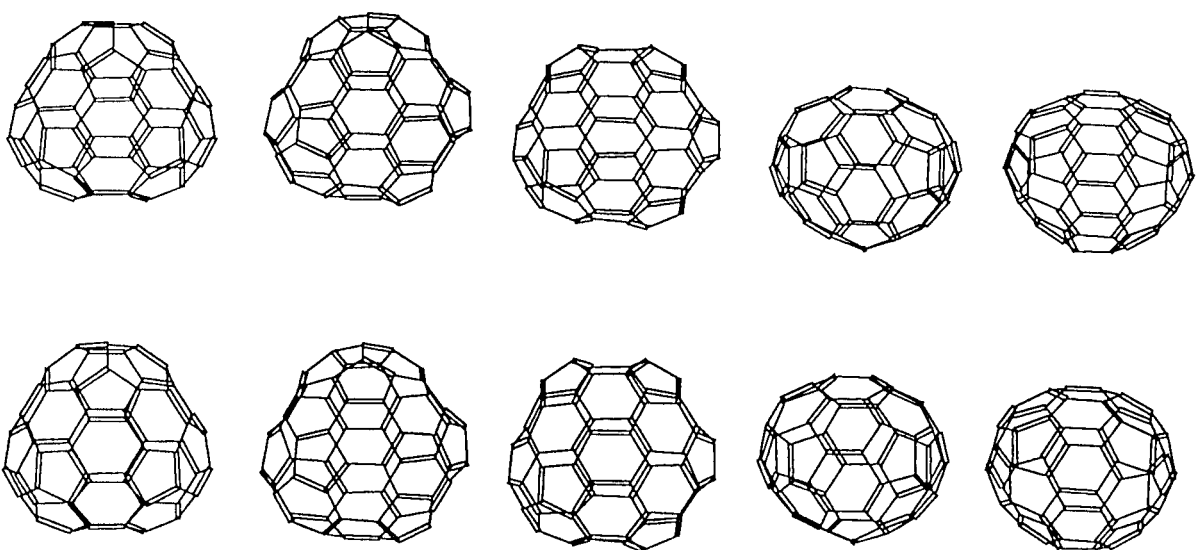


FIG. 3. Stereo-views of symmetrical non-unique C_{60} isomers (From top to bottom): 3A, 3B: $p = 4$, $q = 0$; 3C, 3D, 3E: $p = 6$, $q = 0$.

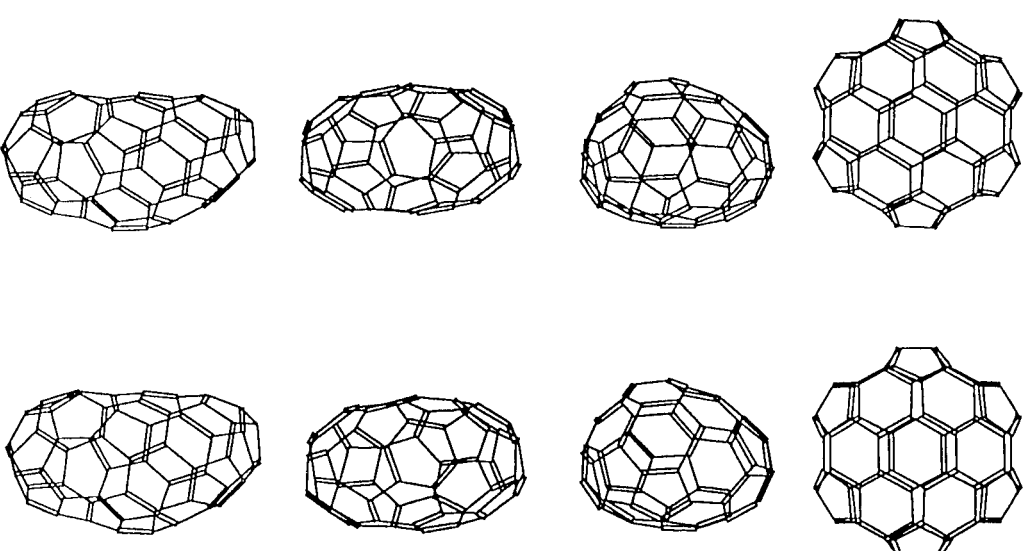


FIG. 3 (continued). Stereo-views of symmetrical C_{60} isomers: 3F, 3G, $p = 6$, $q = 0$; 3H, $p = 8$, $q = 0$; 3I, $p = 10$, $q = 2$.

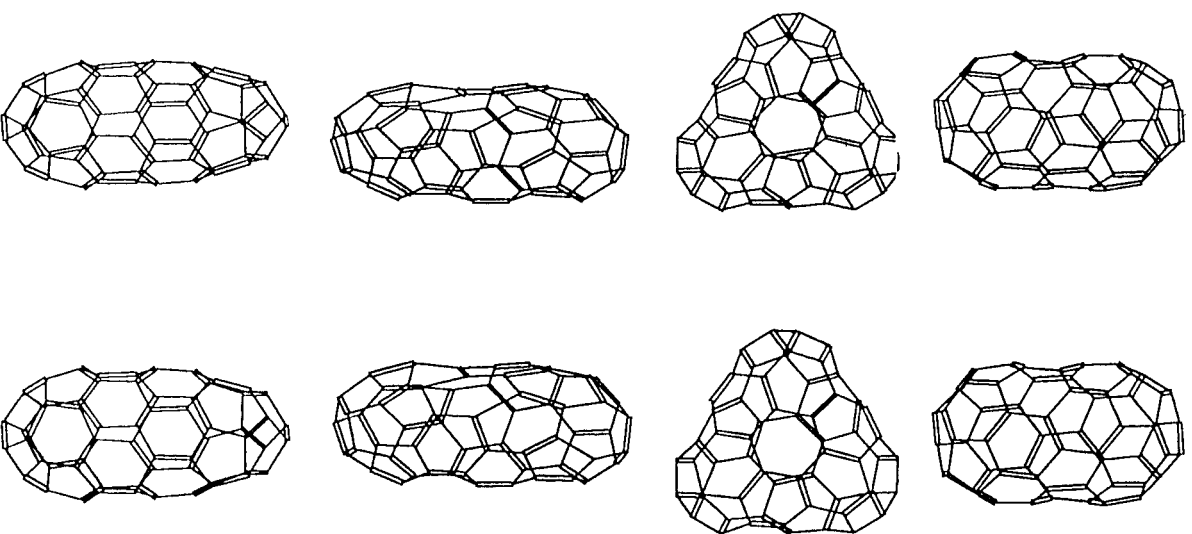


FIG. 3 (continued). Stereo-views of symmetrical C_{60} isomers: **3I**, $p = 10$, $q = 0$; **3K**, $p = 15$, $q = 6$; **3L**, $p = 12$, $q = 2$; **3M**, $p = 14$, $q = 4$.

energy increase. For such cases we introduce the number k of coronene units, i.e. of hexagons surrounded by six hexagons. A further invariant for fullerenes, denoted by r , was recently proposed by Cash;²⁷ it denotes continuous pentagon triplets that have no single vertex in common. It will not be used here. Neither will other invariants enumerated in earlier publications.^{15,17}

The descriptors $a - k, p, q$ do not characterize fullerenes uniquely; they partition the 1812 C_{60} isomers into 622 sets, 345 of which are "unique", i.e. contain one isomer; for the remaining sets, the degeneracy may be fairly high in some cases, up to 33 fullerenes with the same set of descriptors. However, the corresponding energies are within a narrower range (i.e. at most 18 kcal/mol) than the range corresponding to partition *via* p, q invariants.

In Table 4 all molecular descriptors discussed above are presented for the selected C_{60} isomers: the seven "unique" isomers in terms of p and q (upper part of Table 4) and the 12 symmetrical isomers mentioned in a preceding section of the present paper.

Other proper fullerenes

In Table 5 the same data are presented for C_{50} isomers as in Table 3 for the C_{60} isomers, with the same convention (the six unique C_{50} isomers in the terms of p and q are indicated in boldface characters). Similarly to Table 4, Table 6 presents the invariants for the unique C_{50} isomers, three of which are shown in Fig. 4, including the two isomers with extreme energies.

The 271 constitutional isomers of C_{50} are partitioned by the invariants $a - i, k$ into 166 sets, 113 of which are unique; the remaining 53 sets have between two and seven isomers each, with a maximum difference in energy of 12.4 kcal/mol in a set.

Table 4. Parameters p, q, A – D, a – i, k and computed MM energy for selected C₆₀ isomers.

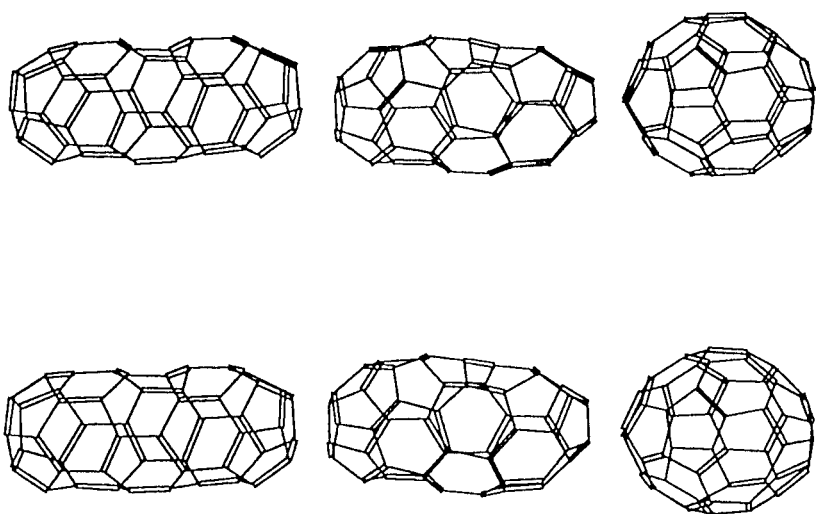
Name	p	q=A	B	C	D	a	b	c	d	e	f	g	h	i	k	Energy	Figure
Buckminsterfullerene	0	0	0	60	0	0	0	0	0	0	60	30	0	0	0	267.63	2A
	2	0	4	52	4	0	0	2	0	8	48	22	8	2	0	282.22	2B
	11	1	19	19	21	0	3	8	8	22	8	1	17	23	2	377.31	2C
Thick-sausage	12	0	24	12	24	0	0	12	12	24	0	0	12	30	6	390.72	2D
	16	6	14	14	26	4	10	2	6	16	6	0	14	32	6	431.22	2E
	18	8	12	12	28	6	12	0	4	16	4	0	12	36	8	454.36	2F
Thin-sausage	20	10	10	10	30	10	10	0	0	20	0	0	10	40	10	477.44	2G
Pillow	4	0	8	44	8	0	0	4	0	16	36	14	16	4	0	297.22	3A
	4	0	8	44	8	0	0	4	0	16	36	14	16	4	0	297.27	3B
	6	0	12	36	12	0	0	6	1	22	25	9	18	9	1	319.46	3C
	6	0	12	36	12	0	0	6	2	20	26	8	20	8	1	319.77	3D
	6	0	12	36	12	0	0	6	3	18	27	9	18	9	1	323.85	3E
	6	0	12	36	12	0	0	6	0	24	24	12	12	12	2	324.22	3F
	6	0	12	36	12	0	0	6	2	20	26	10	16	10	1	324.65	3G
	8	0	16	28	16	0	0	8	4	24	16	4	20	14	0	341.00	3H
	10	2	14	26	18	0	6	4	8	12	20	0	26	14	1	355.77	3I
	10	0	20	20	20	0	0	10	8	24	8	2	16	22	2	366.54	3J
	15	6	12	18	24	3	12	0	0	24	6	0	18	27	2	372.44	3K
	12	2	18	18	22	0	6	6	10	16	10	0	18	24	2	391.17	3L
	14	4	16	16	24	2	8	4	8	16	8	0	16	28	4	417.72	3M

Table 5. MM Energy partition (in kcal/mol) of the 271 C₅₀ isomers according to p and q.

p	q	n(p,q)	E _{ave}	E _{min}	E _{max}	RMS	ΔE
5	0	1	277.13	277.13	277.13	0.000	0.00
6	0	6	285.66	283.35	287.17	1.341	3.82
7	0	11	294.62	291.13	297.81	2.250	6.68
8	0	28	303.02	298.58	310.68	2.703	12.10
8	1	18	299.08	295.14	305.49	2.853	10.35
9	0	12	317.29	311.41	322.72	3.673	11.31
9	1	40	308.37	302.79	316.11	3.340	13.32
9	2	10	304.03	297.01	309.54	3.665	12.53
10	0	4	327.16	319.51	332.26	5.278	12.75
10	1	22	321.29	313.71	331.17	4.187	17.46
10	2	29	313.15	306.47	325.86	4.200	19.39
10	3	5	309.49	304.62	313.86	3.186	9.24
11	1	5	335.69	330.65	339.53	3.436	8.88
11	2	24	326.40	316.04	333.98	4.422	17.94
11	3	10	320.91	312.98	327.03	4.413	14.05
11	4	2	320.27	318.95	321.59	1.320	2.64
12	2	6	340.04	331.05	353.34	7.912	22.29
12	3	9	330.69	317.71	338.58	5.772	20.87
12	4	7	325.96	319.85	333.43	5.420	13.58
13	2	1	343.38	343.38	343.38	0.000	0.00
13	3	2	343.55	341.43	345.68	2.125	4.25
13	4	9	340.72	334.16	347.24	4.419	13.08
13	5	1	343.01	343.01	343.01	0.000	0.00
14	4	2	353.52	347.67	359.37	5.850	11.70
14	5	2	344.70	339.11	350.28	5.585	11.17
14	6	2	351.46	345.75	357.17	5.710	11.42
16	6	1	371.62	371.62	371.62	0.000	0.00
18	8	1	390.45	390.45	390.45	0.000	0.00
20	10	1	402.54	402.54	402.54	0.000	0.00

Table 6. Parameters p, q, A – D, a – i, k and computed MM energy for selected C₅₀ isomers.

p	q=A	B	C	D	a	b	c	d	e	f	g	h	i	k	Energy	Figure
5	0	10	30	10	0	0	5	0	20	30	20	0	0	0	277.13	4A
13	5	11	13	21	3	9	1	3	16	15	4	15	9	0	343.01	-
13	2	20	4	24	0	6	7	10	20	4	0	14	14	0	343.38	-
16	6	14	4	26	4	10	2	6	16	6	0	14	17	1	371.62	4B
18	8	12	2	28	6	12	0	6	12	6	0	12	21	3	390.45	-
20	10	10	0	30	10	10	0	0	20	0	0	10	25	5	402.54	4C

FIG. 4. Three from the six unique C_{50} isomers.

The smallest proper fullerene (with huge strain) is C_{20} . The next unique fullerene is C_{24} . The third one in this series, also a unique isomer, is C_{26} . From the three C_{30} isomers we include in Table 7 the one with D_{5h} symmetry, and from the 15 C_{36} isomers we include the one with D_{6h} symmetry. Two from the 40 isomers of C_{40} , both with D_{5d} symmetry, and two from the 199 isomers of C_{48} , both with D_{6d} symmetry, are also included. Remarkably, both last isomers will be shown to share all descriptors which will be used in our investigation except k . In the present paper we shall discuss in detail C_{30} and C_{60} isomers.

Table 7. A few proper fullerenes with $N < 50$.

N	p	q=A	B	C	D	a	b	c	d	e	f	g	h	i	k	E	Fig.
20(I_h)	30	20	0	0	0	30	0	0	0	0	0	0	0	0	0	189.43	5A
24(D_{6d})	24	12	12	0	0	12	12	0	12	0	0	0	0	0	0	222.05	5B
26(D_{3h})	21	8	18	0	0	6	12	3	18	0	0	0	0	0	0	233.17	5C
30(D_{5h})	20	10	10	10	0	10	10	0	0	20	0	5	0	0	0	254.16	5D
36(D_{6h})	12	0	24	12	0	0	0	12	12	24	0	6	0	0	0	266.12	5E
40(D_{5d})	20	10	10	10	10	10	10	0	0	20	0	0	10	10	0	327.66	5F
40(D_{5d})	10	0	20	20	0	0	0	10	10	20	10	10	0	0	0	284.34	5G
48(D_{6d})	12	0	24	12	12	0	0	12	12	24	0	0	12	12	0	328.54	5H
48(D_{6d})	12	0	24	12	12	0	0	12	12	24	0	0	12	12	2	370.44	5I

In the present paper we describe a set of simple topological invariants allowing to correlate with satisfactory results the energies of proper fullerenes with abutting pentagons.

Correlation with energies of proper fullerenes

Equilibrium geometries and molecular energies were computed by molecular mechanics method, by using MM+ force field as implemented in HyperChem. Initial geometries were also provided by HyperChem.²⁸ In a number of cases, final geometries were examined and checked for possible non-global minimum condition. This was done by perturbing the computed geometry and optimizing it again. Always we obtained convergence to the same minimum. It seems that ellipsoidal structures such as fullerenes do not allow for more than one distinct conformation, as far as molecular mechanics methods are concerned.

While the resulting absolute energies presented in Table 3 (spanning an energy interval of more than 200 kcal/mol) are lower than the E_{r} -energies obtained by Austin et al.¹⁰ using the semiempirical MM QOFF/PI method, the relative energies are in good agreement.

As shown in Table 8, none of the descriptors $a - k$, p and q have individual (partial) correlation factors higher than 0.96 for MM energies. The highest values are obtained for i and p ; i will be selected as one preferred parameter because it does not intercorrelate strongly with any other descriptor than p , followed by f ; both these last descriptors have fairly high intercorrelations with several other descriptors, therefore they will not be selected. The explanation for the strong positive intercorrelation between p and i is that when local agglomerations of pentagons (i.e. high p values) occur, an excess of hexagons (i.e. high i values) remains. It is also possible to explain the negative sign of some intercorrelations⁹ observed in Tables 9 and 10 by the fact that when agglomerations of pentagons

Table 8. Partial correlation coefficients with MM energies.

Parameter	C_{60}	C_{50}	$C_{50} + C_{60}$
a	0.576	0.680	0.502
b	0.593	0.641	0.519
c	-0.103	-0.339	-0.147
d	0.596	0.460	0.447
e	0.111	-0.409	0.014
f	-0.850	-0.808	-0.650
g	-0.720	-0.804	-0.702
h	-0.266	0.387	0.051
i	0.956	0.960	0.946
k	0.751	0.500	0.733
p	0.924	0.939	0.744
q	0.654	0.734	0.573
N			0.315

(high a , b , p , q) or hexagons (high i , k) occur, which both lead to strain and energy increase, then quasi-spherical low-strain structures $c - h$ must remain fewer in number. In other words, in the parameter series starting with agglomerations of pentagons and ending with the agglomeration of hexagons (p , q , $a - i$, k), parameters at extremities are positively intercorrelated among themselves, and negatively intercorrelated with those in the middle. Since p is strongly intercorrelated with i and q , as seen in Tables 9 and 10, we shall also select q as a second preferred parameter.

Linear correlations

(i) The 1812 C_{60} isomers.

For the C_{60} isomers the following equations hold:

$$C = 60 - 3A - 2B = 60 - 4p + 3q \quad (10)$$

$$D = 2A + B = 2p - q \quad (11)$$

Table 9. Intercorrelation matrix for the parameters a - i, k, p, q for C₆₀.

	b	c	d	e	f	g	h	i	k	p	q
a	0.547	-0.455	-0.075	-0.076	-0.365	-0.295	-0.208	0.553	0.595	0.639	0.761
b		-0.820	-0.059	-0.213	-0.401	-0.402	-0.127	0.630	0.424	0.782	0.959
c			0.507	0.487	-0.173	-0.123	0.130	-0.128	-0.153	-0.318	-0.789
d				-0.033	-0.562	-0.564	-0.041	0.523	0.342	0.397	-0.071
e					-0.554	-0.408	0.167	0.216	-0.033	0.156	-0.191
f						0.887	0.014	-0.890	-0.515	-0.859	-0.434
g							-0.409	-0.694	-0.345	-0.788	-0.411
h								-0.357	-0.411	-0.117	-0.169
i									0.695	0.945	0.675
k										0.638	0.529
p											0.822

Table 10. Intercorrelation matrix for the parameters a - i, k, p, q for C₅₀.

	b	c	d	e	f	g	h	i	k	p	q
a	0.589	-0.604	-0.206	-0.318	-0.373	-0.352	0.070	0.685	0.711	0.745	0.850
b		-0.895	-0.096	-0.526	-0.367	-0.532	0.402	0.633	0.255	0.820	0.926
c			0.460	0.518	-0.034	0.139	-0.130	-0.339	-0.257	-0.559	-0.865
d				-0.314	-0.505	-0.547	0.318	0.391	-0.062	0.247	-0.159
e					-0.113	0.146	-0.038	-0.319	-0.165	-0.384	-0.491
f						0.903	-0.565	-0.821	-0.294	-0.776	-0.413
g							-0.822	-0.734	-0.200	-0.821	-0.511
h								0.247	-0.085	0.490	0.295
i									0.515	0.929	0.731
k										0.475	0.498
p											0.882

$$g = 30 + p - i - h \quad (12)$$

$$h = (3q - 6g - 4i + 180)/5 \quad (13)$$

$$p = (h + 2i + 3q)/6 \quad (14)$$

$$q = A = (2a - 2c - 2d - e + h + 2i)/9 \quad (15)$$

For the 1812 isomers of C₆₀, six, five, or four, parameters were used paying attention to have between these parameters intercorrelation factors lower than 0.7. The results are presented in Table 11(i).

A remarkably good agreement was found between energies calculated by the MM method and from these linear correlations.

The first correlation with six parameters has only three outliers with more than 15 kcal/mol. We shall denote by OL the number of isomers with absolute residual values greater than a specified threshold, and by *s* the standard deviation. The next four correlations in terms of five parameters have statistical parameters which are comparable to those of the first correlation including outliers OL > 3*s*, except for the number of outliers in the last column, therefore we consider that five parameters offer the best solutions. Already with four parameters (the last correlation) the statistical deviation *s* and the number of outliers OL > 3*s* are worse than for correlations with five parameters.

Among the five 5-parameter correlations, we prefer the last three ones because they have the best statistical parameters. Because *g* and *h* are linearly interdependent (eq. 12, 13) without involving *d* and *k*, these last two 5-parameter correlations have exactly the same statistical parameters (last five columns) and the values of coefficients for *d* and *k* are the same. It will be seen that this is a general feature in Table 11(i). The correlations including parameter *a* have worse statistical parameters (*s* and OL > 15), and again this is a general feature of Table 11(i). The parameters *c* and *q* have a fairly high intercorrelation coefficient (-0.79); we have avoided the use of parameters that have intercorrelation coefficients higher than 0.80.

Table 11. Regression coefficients and, statistical indices r , s and F for multilinear regression models relating MM energies to selected topological descriptors for fullerenes. The tables contain coefficients of parameters, the free term (intercept, in kcal/mol), and the number of outliers exceeding either $3s$, or 15 kcal/mol.

(i) The 1812 C_{60} fullerene isomers														
intercept	a	c	d	g	h	i	k	q	N	r	s	F	OL>3 s	OL>15 kcal/mol
278.920	2.748		2.502		0.532	2.728	3.900	1.829		0.980	4.015	7169	14	3
290.788	2.960		2.922			2.495	3.207	2.465		0.978	4.175	7926	15	7
267.152	1.295				0.984	3.745	5.054	-0.872		0.971	4.776	5973	14	12
279.152		2.976	2.543			1.719	3.798	9.254		0.980	4.015	8601	15	3
298.236			2.319	-0.673		2.309	4.638	2.871		0.978	4.127	8120	15	7
278.054			2.319		0.561	2.757	4.638	2.535		0.978	4.127	8121	15	7
290.544			2.749			2.514	3.965	3.266		0.977	4.301	9310	17	7
(ii) The 271 C_{50} fullerene isomers														
285.358	2.632		2.656			2.532	0.187	2.988		0.980	3.292	1312	3	0
278.971	1.579				1.119	4.219	0.974	-0.607		0.974	3.775	985	2	0
263.210		3.362	2.209			1.640	0.301	10.144		0.984	3.007	1584	3	0
293.597			1.834	-0.694		2.567	3.395	3.131		0.981	3.223	1371	3	0
279.725			1.834		0.578	3.030	3.395	2.784		0.981	3.223	1371	3	0
(iii) The 271 C_{50} plus 1812 C_{60} fullerene isomers														
256.277	2.702		2.877			2.502	3.174	2.575	0.575	0.980	4.088	8472	15	5
367.660	1.416				1.038	3.792	4.777	-0.794	-1.703	0.974	4.698	6330	17	12
193.059		2.878	2.496			1.759	3.730	9.026	1.451	0.981	3.918	9252	21	5
267.233			2.269	-0.684		2.313	4.565	2.991	0.518	0.981	4.037	8696	17	4
287.750			2.269		0.570	2.769	4.565	2.649	-0.166	0.981	4.037	8696	17	4

(i) The 271 C_{50} isomers.

A similar approach led for the C_{50} isomers to the four 5-parameter correlations presented in Table 11(ii). The same observations apply as indicated above, and the statistical parameters (r , s , OL) are better than for the C_{60} isomers. For this set, smaller than the preceding one, 4-parameter correlations are also possible with comparable statistical parameters. However, they have not been included.

(iii) The combined isomers of C_{50} and C_{60} .

As seen in the four 6 - parameter correlations shown in the Table 11(iii) with N as an additional descriptor, the statistical data are equal to, or intermediate between, those for the two separate sets. The correlation coefficient r is equal to, or higher than, the best one in Tables 11(i) and 11(ii), only s and $OL > 15$ are intermediate. For the last two best correlations, one of the coefficients (g or N) has a negative sign.

A special mention should be made of the only serious outlier: the "thick sausage" for the C_{60} isomers, according to the multilinear regression, should have a lower energy than that obtained by the MM calculations by about 30 kcal/mol; we had suspected that a non-global minimum was obtained in the calculations, but this does not seem to be the case.

(iv) Correlations including also proper fullerenes with $N < 50$

We selected the 35 fullerenes included in Tables 4 (twenty C_{60} isomers), 6 (six C_{50} isomers) and 7 (nine fullerenes with $N < 50$). For this set, a 5-parameter correlation was found:

$$E = 176.135 + 2.134b + 1.336e - 0.141g + 3.732i + 1.501N$$

$$n = 35 \quad r = 0.969 \quad s = 17.39 \text{ kcal/mol}$$

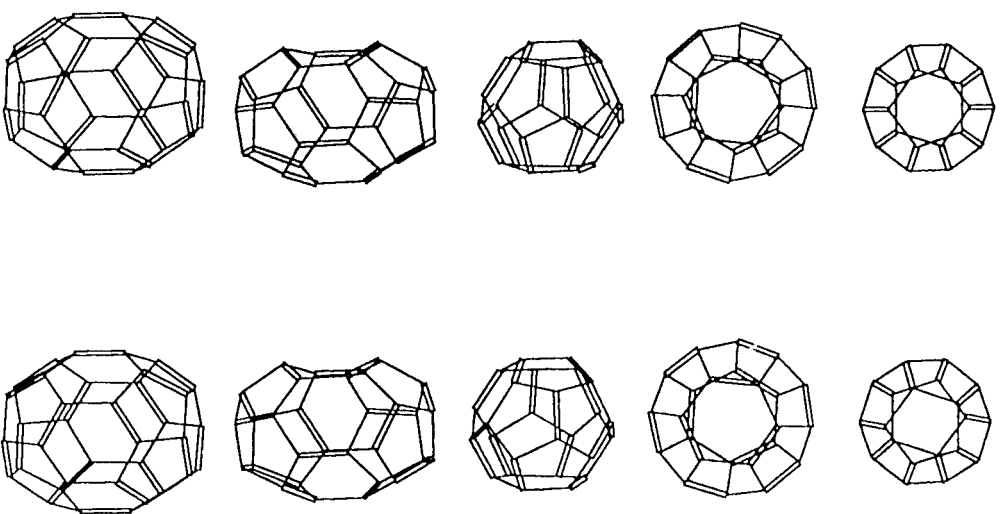


FIG 5. Stereo-views proper fullerenes with $N < 50$ (from top to bottom): the unique isomers (in terms of p, q -parameters) 5A, C_{20} , 5B, C_{24} , 5C, C_{26} ; next, non-unique isomers: 5D, C_{30} (D_{5h}); 5E, C_{36} (D_{6h}).

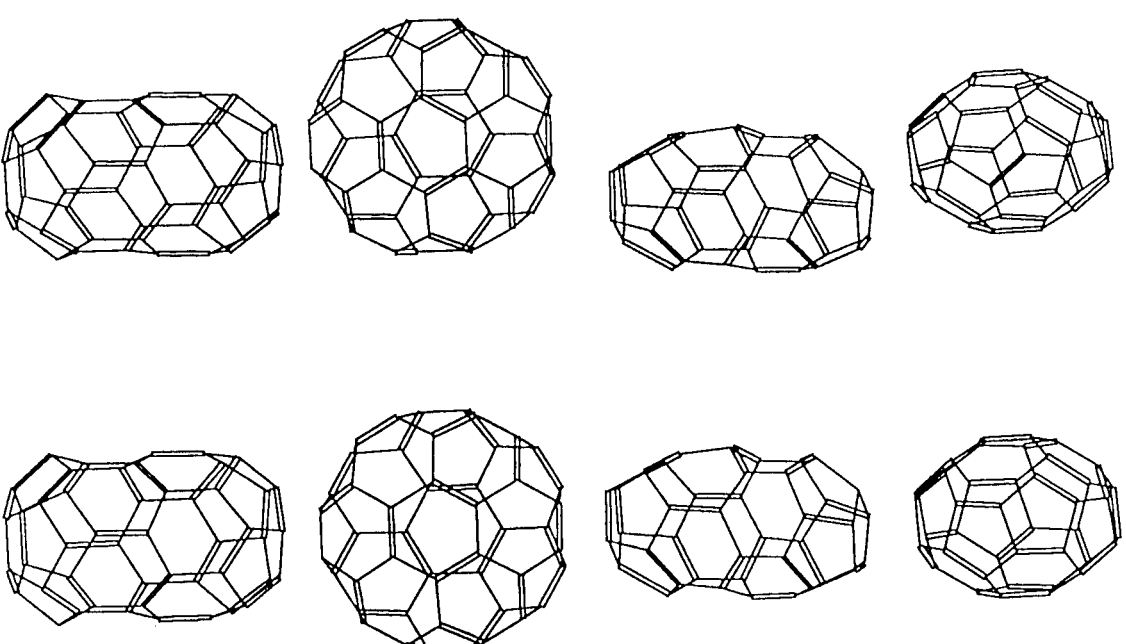


FIG 5 (continued). Stereo-views proper fullerenes with $N < 50$ (from top to bottom): the unique isomers (in terms of p, q -parameters) 5F, C_{40} (D_{3d}) with $p = 20$, $q = A = 10$; 5G, C_{40} (D_{3d}) with $p = 10$, $q = A = 0$; 5H, C_{48} (D_{6d}) with $p = 12$, $q = A = 0$, $k = 0$; 5I, C_{48} (D_{6d}) with $p = 12$, $q = A = 0$, $k = 2$.

Three serious outliers are the C₄₀ (D_{5d}) isomer 5G, the C₄₈ (D_{6d}) isomer 5I (which is analogous to the "thick sausage" isomer of C₆₀), and the C₆₀ isomer 3K. A possible explanation resides in the fact that the topological invariants selected in our correlation are good for characterizing the *curvature*, but not the *angle* and *bond* deviations. If these three isomers are omitted, the correlation improves markedly, with a standard deviation which is lower than that found earlier by Austin et al.:²²

$$E = 144.806 + 3.788 b + 1.432 e + 0.063 g + 3.546 i + 1.945 N$$

$$n = 32 \quad r = 0.992 \quad s = 9.08 \text{ kcal/mol}$$

Conclusions

For the energies of the 1812 constitutional isomers of C₆₀ and the 271 isomers of C₅₀, which span a range differing by more than 200 kcal/mol, multilinear correlations were developed in terms of molecular descriptors consisting in the topological environment of each edge. For the combined set of 1812 + 271 isomers, including also the number N of carbon atoms as a parameter, such correlations afford satisfactory results: correlation coefficient r around 0.98, standard deviation s around 4 kcal/mol. For a small set of only 35 proper fullerenes with 20 to 60 carbon atoms the analogous correlations give much poorer results, therefore such systems need further investigations.

Acknowledgement: One of the authors (OJ) acknowledges a software grant consisting of HyperChem Release 4.5 offered by Hypercube, Inc.

References

- [1] Y.D. Gao and W.C. Herndon, *J. Am. Chem. Soc.*, 1993, **115**, 8459.
- [2] D. Babic and N. Trinajstić, *Chem. Phys. Lett.*, 1995, **237**, 239.
- ENERGY AND TOPOLOGICAL INVARIANT CORRELATION. I 1505
- [3] D. Babic, A. Graovac and N. Trinajstić, *Croat. Chem. Acta*, 1993, **66**, 35.
- [4] M. Randić, *Fullerene Sci. Technol.*, 1994, **2**, 427.
- [5] R.L. Murry, D.L. Strout, G.K. Odom and G.E. Scuseria, *Nature*, 1993, **366**, 665.
- [6] H.W. Kroto, J.R. Heath, S.C. O'Brien, R.F. Curl and R.E. Smalley, *Nature*, 1985, **318**, 162.
- [7] X. Liu, D.J. Klein, T.G. Schmalz and W.A. Seitz, *J. Comput. Chem.*, 1991, **12**, 1252.
- [8] X. Liu, D.J. Klein, W.A. Seitz and T.G. Schmalz, *J. Comput. Chem.*, 1991, **12**, 1265.
- [9] T.G. Schmalz, D.J. Klein and X. Liu, *Mat. Res. Soc. Symp. Proc.*, 1992, **270**, 129.
- [10] T.G. Schmalz, X. Liu and D.J. Klein, *Chem. Phys. Lett.*, 1992, **192**, 331.
- [11] D.E. Manolopoulos, *Chem. Phys. Lett.*, 1992, **192**, 330.
- [12] P.W. Fowler, *Contemporary Phys.*, 1996, **37**, 235.
- [13] H.W. Kroto, *Nature*, 1987, **329**, 529.
- [14] T.G. Schmalz, W.A. Seitz, D.J. Klein and G.E. Hite, *J. Am. Chem. Soc.*, 1988, **110**, 1113.
- [15] B.L. Zhang, C.Z. Wang, K.M. Ho, C.H. Xu and C.T. Chan, *J. Chem. Phys.*, 1992, **97**, 5007.
- [16] B.L. Zhang, C.Z. Wang, K.M. Ho, C.H. Xu and C.T. Chan, *J. Chem. Phys.*, 1993, **98**, 3095.
- [17] A.T. Balaban, X. Liu, D.J. Klein, D. Babic, T.G. Schmalz, W.A. Seitz and M. Randić, *J. Chem. Inf. Comput. Sci.*, 1995, **35**, 396.
- [18] A.T. Balaban, D. Babic and D.J. Klein, *J. Chem. Educ.*, 1995, **72**, 693.
- [19] D.J. Klein and H. Zhu, in "From Chemical Topology to Three-Dimensional Geometry" (A.T. Balaban, Ed.), Plenum Press, New York, 319.

- [20] A.T. Balaban, T.G. Schmalz, H. Zhu and D.J. Klein, *J. Mol. Struct. Theochem*, 1996, **363**, 291.
- [21] S.J. Austin, P.W. Fowler, G. Orlandi, D.E. Manolopoulos and F. Zerbetto, *Chem. Phys. Lett.*, 1994, **226**, 219.
- [22] S.J. Austin, P.W. Fowler, D.E. Manolopoulos, G. Orlandi and F. Zerbetto, *J. Phys. Chem.*, 1995, **99**, 8076.
- [23] P.W. Fowler and D.E. Manolopoulos, *An Atlas of Fullerenes*, Oxford Univ. Press, 1994.
- [24] A.J. Stone and D.J. Wales, *Chem. Phys. Lett.*, 1986, **128**, 501.
- [25] A.T. Balaban, T.G. Schwalz, H. Zhu and D.J. Klein, *Theochem* 1996, **363**, 291.
- [26] D. Babic, S. Bassoli, M. Casartelli, F. Cataldo, A. Graovac, O. Ori and B. York, *Mol. Simul.* 1995, **14**, 395.
- [27] G.C. Cash, *Polycyclic Arom. Compsd.*, 1997, **12**, 61.
- [28] HyperChem Release 4.5 for Windows, Hypercube, Inc., 419 Phillip St., Waterloo, Ontario, Canada N2L 3X2; Tel. (519)-725-4040; Fax: (519)-725-5193; information hot-line: (800)-960-1871; E-mail information requests: info@hyper.com; URL: <http://www.hyper.com>.

(Received June 3, 1997)