

SI: Mapping Structure-Property Relationships in Fullerene Systems: A Computational Study from C₂₀ to C₆₀

Bin Liu^{1,2}, Jirui Jin^{1,2}, and Mingjie Liu^{1,2*}

¹Department of Chemistry, University of Florida, Gainesville, FL 32611, United States.

²Quantum Project Theory, University of Florida, Gainesville, FL 32611, United States.

*mingjieliu@ufl.edu

Table S1: Statistic analysis of binding energy (E_b (eV/atom)), HOMO-LUMO gap (E_g (eV)), dipole moment (μ (D)), solvation free energies (kJ/mol) in water ($\Delta G_{sol}(water)$) and ODCB ($\Delta G_{sol}(ODCB)$), ODCB-water partition coefficient logP.

	10%	90%	mean	std	max	max-#	min	min-#
E_b	-6.849	-6.722	-6.79	0.056	-6.119	C ₂₀ -#1	-6.958	C ₆₀ -#1
E_g	0.967	1.543	1.243	0.23	2.722	C ₆₀ -#1	0.408	C ₆₀ -#1748
μ	0.235	1.589	0.857	0.545	3.975	C ₆₀ -#1646	0	C ₂₈ -#1
$\Delta G_{sol}(water)$	-48.807	-42.099	-45.735	2.884	-21.316	C ₂₀ -#1	-67.739	C ₅₆ -#3
$\Delta G_{sol}(ODCB)$	-162.08	-139.904	-153.08	10.014	-76.209	C ₂₀ -#1	-168.406	C ₅₆ -#3
logP	17.099	19.993	18.795	1.321	20.223	C ₆₀ -#1790	9.611	C ₂₀ -#1

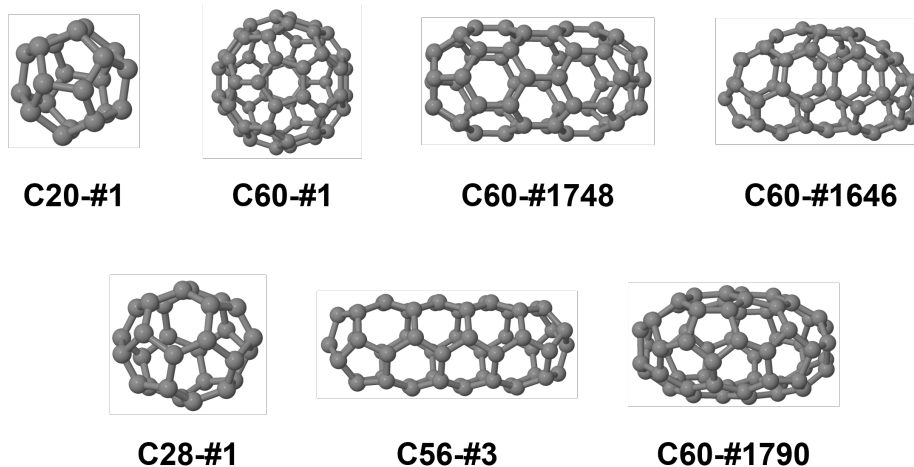


Figure S1: Geometries of the seven fullerene isomers as presented in Table S1.

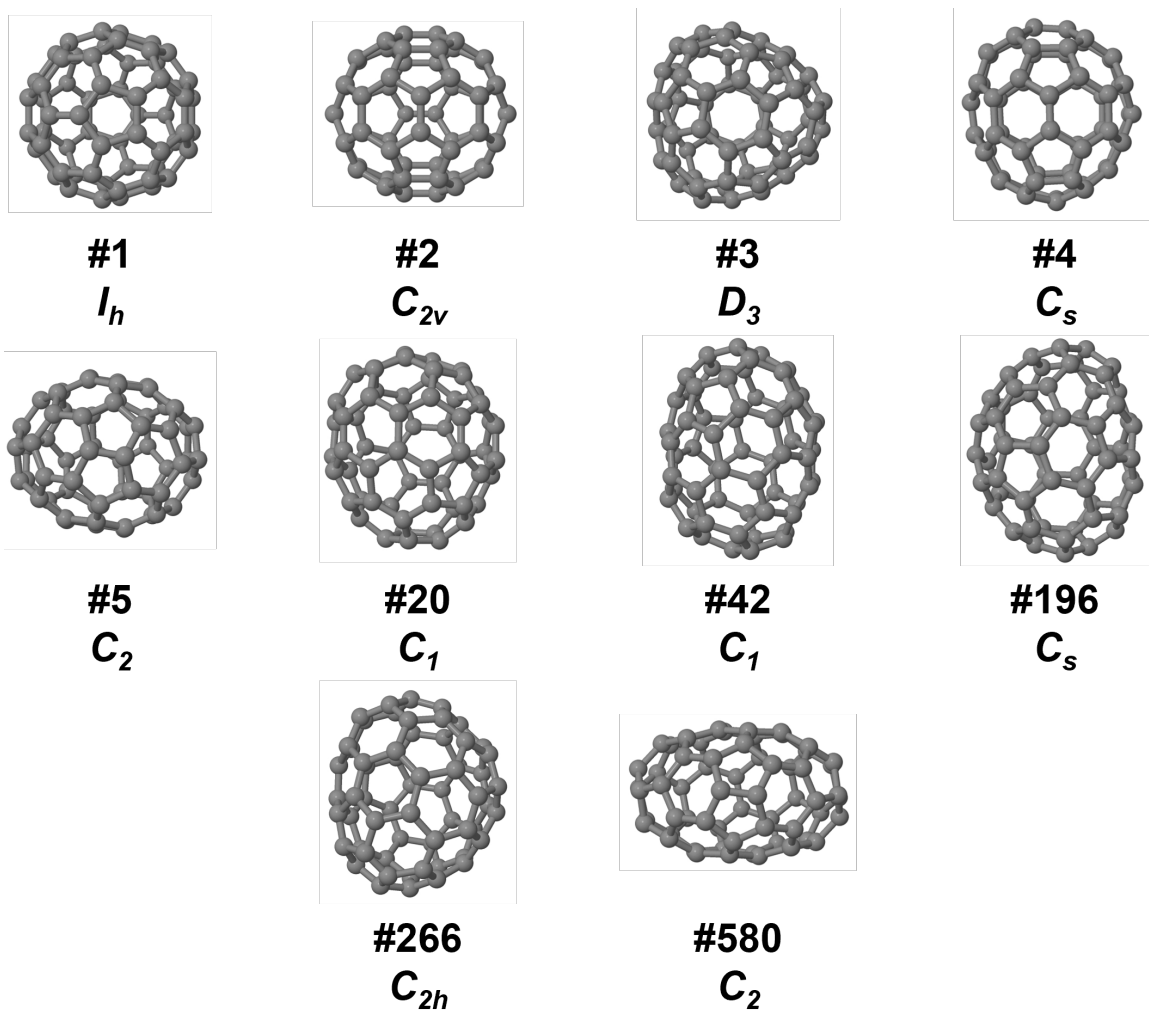


Figure S2: Geometries of the ten C_{60} isomers and their corresponding symmetries.

Table S2: Relative energies (kcal/mol) w.r.t. C_{60} -#1 at the B3LYP-D3/6-311G* (from this work) compared to the DLPNO-CCSD(T)/CBS* reference values adopted from reference[1] and various statistical measures, i.e., the mean deviation (MD), the overall mean absolute deviation (MAD), the standard deviation (SD), the maximal absolute error (MAX). Energies, MD, MAD, SD, and MAX are given in kcal/mol. Geometries of the ten C_{60} isomers are presented in Figure S2

#isomer	Relative energies (kcal/mol) w.r.t to C_{60} -#1	
	DLPNO-CCSD(T)/CBS*	B3LYP-D3/6-311G*
2	37.4	39.05
3	56.87	58.43
4	56.89	58.16
5	69.89	71.29
20	100.48	96.2
42	111.72	111.3
196	142.18	137.05
266	143.96	140.84
580	164.88	165.26
	MD	-0.7
	MAD	2.3
	SD	2.5
	MAX	5.1

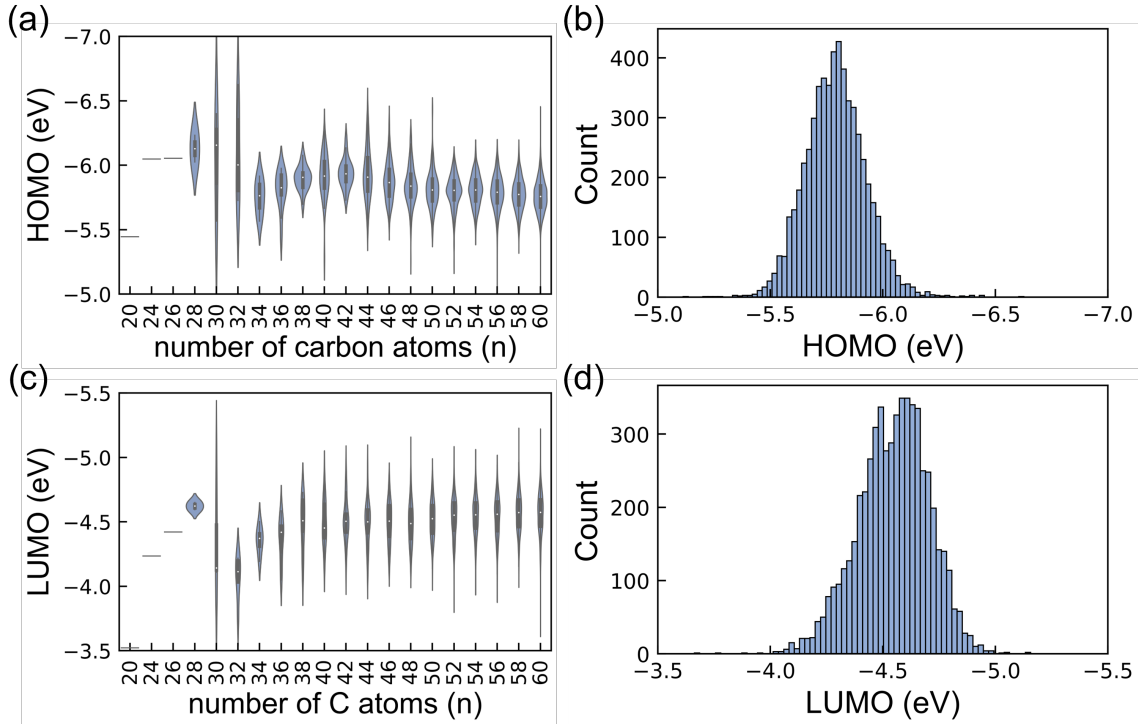


Figure S3: Violin plot and histogram plot for (a,b) HOMO level and (c,d) LUMO level distribution for C_{20} - C_{60} fullerenes.

Table S3: Relative binding energies ΔE_b (kcal/mol) w.r.t. C_{60} -#1 at the B3LYP-D3/6-311G* compared to the DLPNO-CCSD(T)/CBS* reference values and various statistical measures, i.e., the mean deviation (MD), the overall mean absolute deviation (MAD), the standard deviation (SD), the maximal absolute error (MAX). Energies, MD, MAD, SD, and MAX are given in kcal/mol.

#isomer	Relative binding energies ΔE_b (kcal/mol) w.r.t to C_{60} -#1	
	DLPNO-CCSD(T)/CBS*	B3LYP-D3/6-311G*
C_{20} -#1	19.06	19.35
C_{24} -#1	16.17	16.47
C_{26} -#1	14.83	14.92
C_{28} -#1	13.28	13.22
C_{30} -#1	12.99	12.95
C_{32} -#1	10.79	10.69
C_{34} -#1	11.06	10.99
C_{36} -#1	9.87	9.9
C_{38} -#1	9.65	9.64
C_{40} -#1	11.2	11.04
C_{42} -#1	8.52	8.43
C_{44} -#1	8.46	8.39
C_{46} -#1	8.2	8.15
C_{48} -#1	7.93	7.78
C_{50} -#1	9.53	9.57
C_{52} -#1	7.18	7.11
C_{54} -#1	7.13	6.98
C_{56} -#1	6.88	6.84
C_{58} -#1	6.72	6.61
	MD	-0.02
	MAD	0.10
	SD	0.08
	MAX	0.30

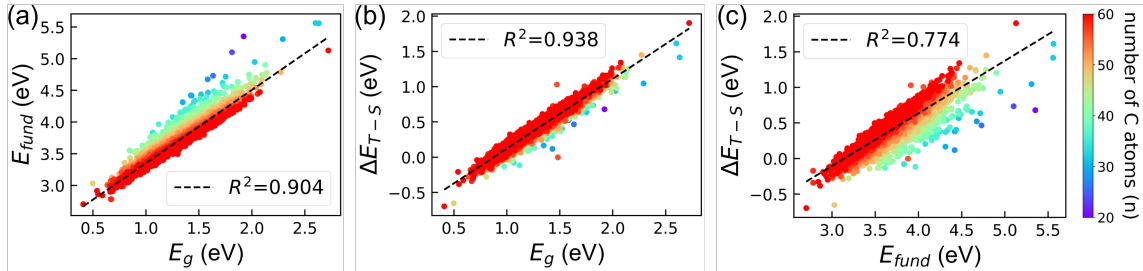


Figure S4: Correlation between HOMO-LUMO gap (E_g (eV)), IP-EA gap (E_{fund} (eV)), and triplet-singlet energy difference (ΔE_{T-S} (eV)) for C_{20} - C_{60} fullerenes.

Table S4: Squared correlation coefficients (R^2) among HOMO-LUMO gap (E_g), fundamental gap (E_{fund}), and triplet-singlet energy differences (ΔE_{T-S}) for C₂₀–C₆₀ fullerenes.

#C	# isomers	E_g/E_{fund}	$E_g/\Delta E_{T-S}$	$E_{fund}/\Delta E_{T-S}$
20	1	1	1	1
24	1	1	1	1
26	1	1	1	1
28	2	1	1	1
30	3	0.999	0.973	0.981
32	6	0.997	0.973	0.971
34	6	0.996	0.986	0.984
36	15	0.966	0.911	0.852
38	17	0.977	0.958	0.969
40	40	0.985	0.962	0.96
42	45	0.983	0.961	0.956
44	89	0.991	0.979	0.97
46	116	0.987	0.964	0.955
48	199	0.99	0.971	0.964
50	271	0.984	0.971	0.964
52	437	0.982	0.955	0.952
54	580	0.983	0.954	0.948
56	924	0.985	0.951	0.944
58	1205	0.984	0.955	0.949
60	1812	0.987	0.964	0.957
All	5770	0.904	0.938	0.774

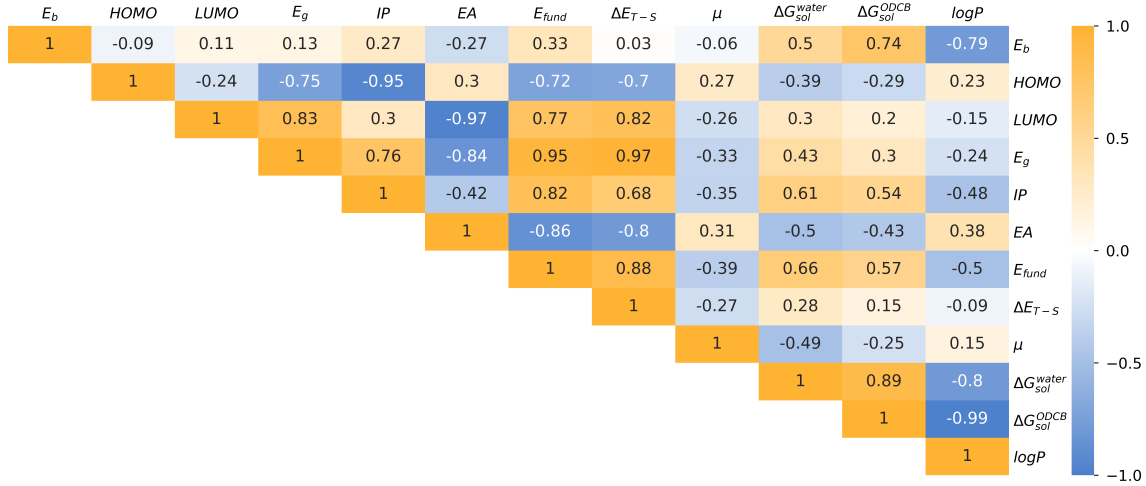


Figure S5: Pearson correlation coefficients among 12 fundamental properties for C₂₀–C₆₀ fullerenes.

Table S5: Coefficient of determination R^2 obtained from linear fitting of topological features and geometrical measures with fundamental properties for the 1812 C_{60} database.

Topological features	R^2		
	E_b	E_g	logP
Atom features ($\{n_i i = 0, 1, 2, 3\}$)	0.92	-0.006	0.297
Bond features ($\{e_i i = 0, \dots, 8\}$)	0.924	0.052	0.353
Pentagon features ($\{p_i i = 0, \dots, 7\}$)	0.924	0.045	0.402
Hexagon features ($\{h_i i = 0, \dots, 12\}$)	0.956	0.143	0.488
Geometrical measures			
Average bond length ($\bar{d}$)	0.901	0.001	0.053
Fowler asymmetry parameter (F_{asym})	0.714	0.002	0.392
Average Curvature ($\bar{\kappa}$)	0.904	0.003	0.314
Volume (V)	0.85	0.004	0.32
Surface area (A)	0.627	0	0.064
Ratio A/V	0.854	0.004	0.32
Deviation from isoperimetric quotient (D_{IPQ})	0.875	0.003	0.3

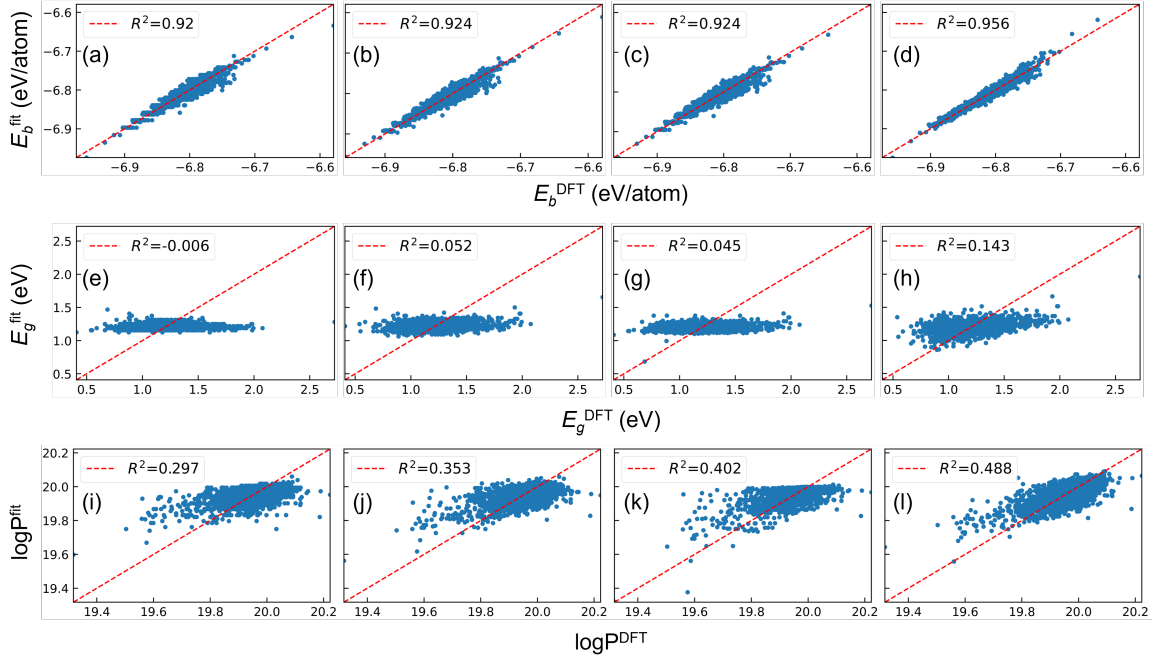


Figure S6: The parity plot of DFT calculated binding energy (E_b (eV/atom)) (left panel), HOMO-LUMO gap (E_g (eV)) (middle panel), and ODCB-water partition coefficient (logP) (right panel) of 1812 C_{60} fullerenes versus predicted values of linear regression based on (a,b,c) atom features, (d,e,f) bond features, (g,h,i) pentagon features, and (j,k,l) hexagon features.

Table S6: The values of 4 topological features, i.e., atom, bond, pentagon, and hexagon features, for four larger-size fullerenes (Cn), i.e., C₇₀, C₈₀, C₉₀, and C₁₀₀. The molecular geometries are presented in Figure S7.

	Cn	C ₇₀	C ₈₀	C ₉₀	C ₁₀₀
Atom features	n0	10	20	30	40
	n1	60	60	60	60
	n2	0	0	0	0
	n3	0	0	0	0
Bond features	e0	5	20	35	50
	e1	60	60	60	60
	e2	20	20	20	20
	e3	0	0	0	0
	e4	0	0	0	0
	e5	20	20	20	20
	e6	0	0	0	0
	e7	0	0	0	0
	e8	0	0	0	0
Pentagon features	p0	12	12	12	12
	p1	0	0	0	0
	p2	0	0	0	0
	p3	0	0	0	0
	p4	0	0	0	0
	p5	0	0	0	0
	p6	0	0	0	0
	p7	0	0	0	0
Hexagon features	h0	0	0	0	0
	h1	0	0	0	0
	h2	0	0	0	0
	h3	0	0	0	0
	h4	0	0	0	0
	h5	0	0	0	0
	h6	0	0	0	0
	h7	10	10	10	10
	h8	0	0	0	0
	h9	10	10	10	10
	h10	5	0	0	0
	h11	0	10	10	10
	h12	0	0	5	10

Table S7: Coefficient of determination R^2 obtained from linear fitting between topological indices and fundamental properties for the 5770 C₂₀–C₆₀ database and the 1812 C₆₀ database.

	R^2 for all C ₂₀ –C ₆₀			R^2 for C ₆₀		
Topological indices	E_b	E_g	logP	E_b	E_g	logP
1 st moment pentagon signature (P_1)	0.821	0.008	0.271	0.912	0.008	0.267
1 st moment hexagon signature (H_1)	0.33	0.066	0.896	0.9	0.007	0.277
2 nd moment hexagon signature (H_2)	0.164	0.059	0.747	0.922	0.007	0.294

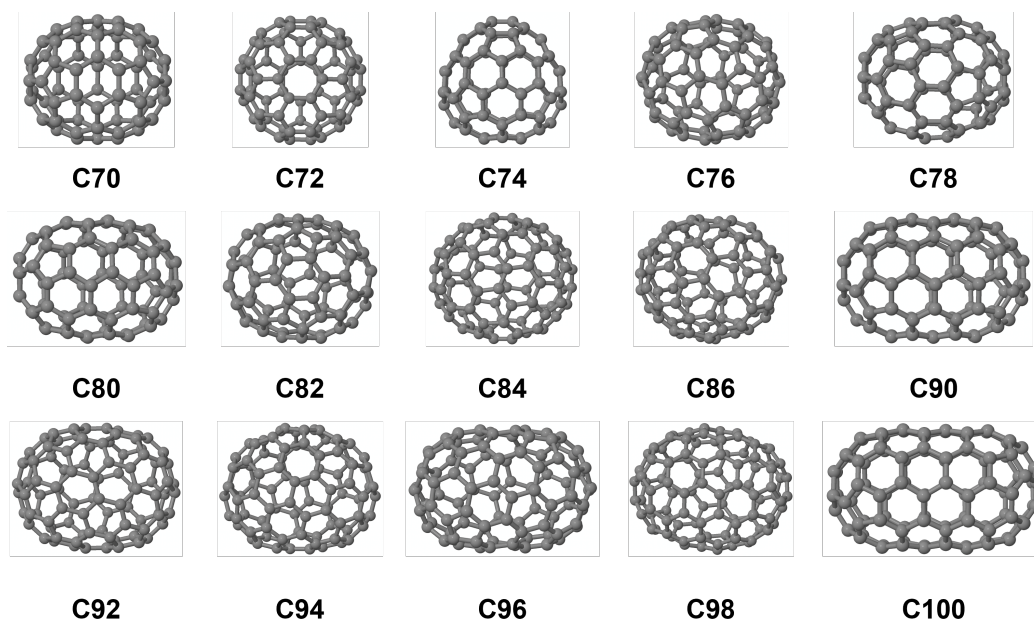


Figure S7: Geometries of 15 large-size fullerenes.

Table S8: The number of possible isomers, hexagon rings, carbon-carbon bonds, and IPR (isolated pentagon rule) isomers for each fullerene cage size

#C	#face	#hexagon	#edge	#isomer	#IPR
20	12	0	30	1	0
22	13	1	33	0	0
24	14	2	36	1	0
26	15	3	39	1	0
28	16	4	42	2	0
30	17	5	45	3	0
32	18	6	48	6	0
34	19	7	51	6	0
36	20	8	54	15	0
38	21	9	57	17	0
40	22	10	60	40	0
42	23	11	63	45	0
44	24	12	66	89	0
46	25	13	69	116	0
48	26	14	72	199	0
50	27	15	75	271	0
52	28	16	78	437	0
54	29	17	81	580	0
56	30	18	84	924	0
58	31	19	87	1205	0
60	32	20	90	1812	1
sum				5770	1

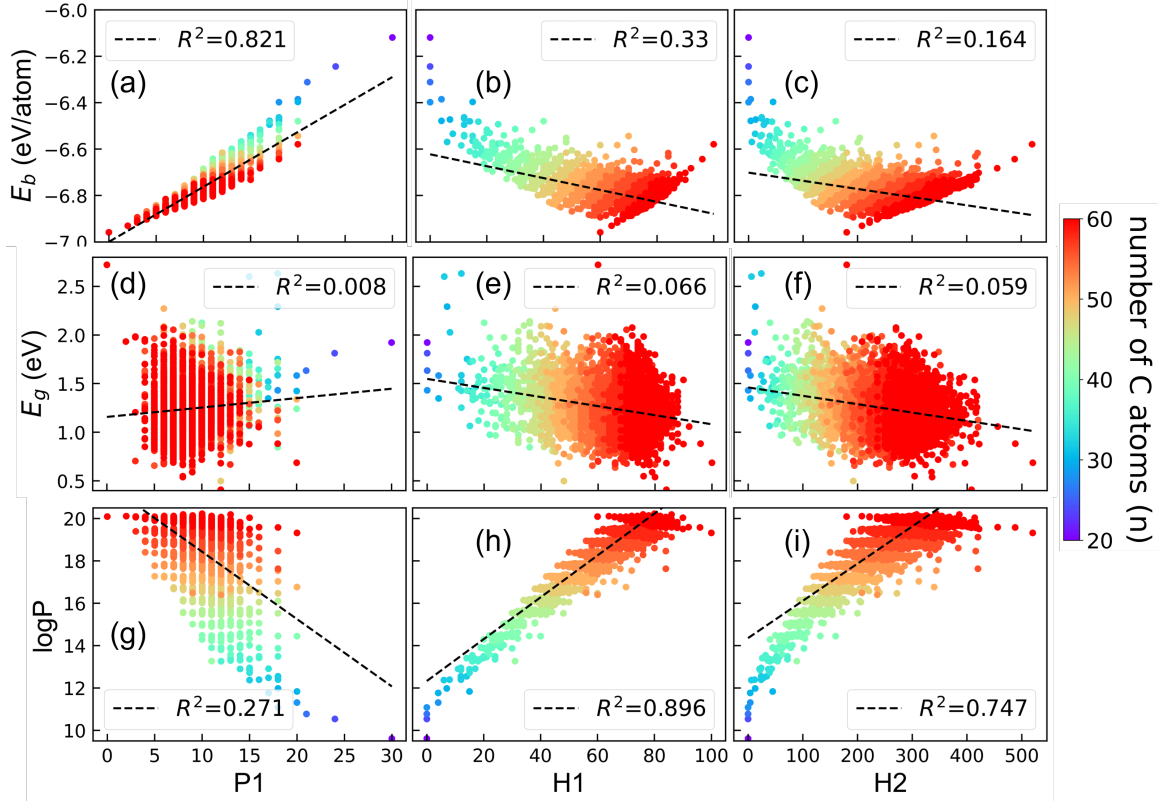


Figure S8: (a-c) The binding energy (E_b (eV/atom)), (d-f) HOMO-LUMO gap (E_g (eV)), and (g-i) ODCB-water partition coefficient ($\log P$) of the whole C_{20} - C_{60} database as a function of the first moment pentagon signature P1 (left panel), the first moment hexagon signature H1 (middle panel), and the second moment hexagon signature H2 (right panel). The dashed lines correspond to a linear fit of the C_{20} - C_{60} dataset. The color bar indicates varying fullerene sizes.

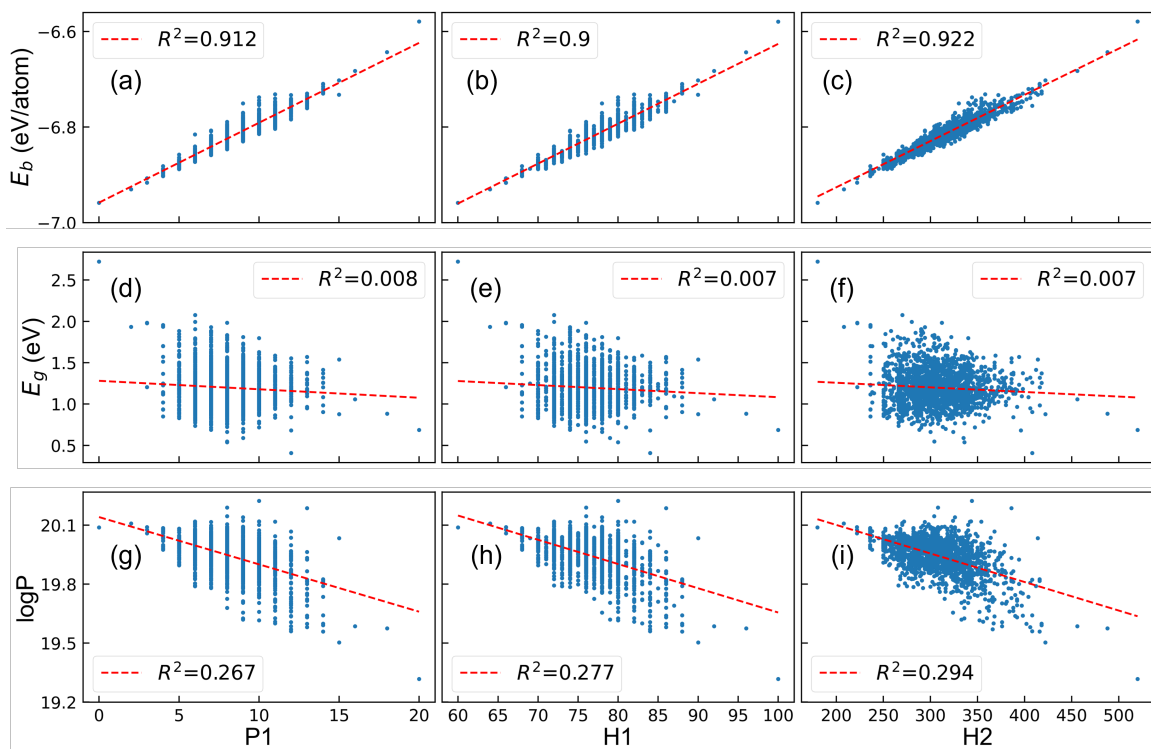


Figure S9: (a-c) The binding energy (E_b (eV/atom)), (d-f) HOMO-LUMO gap (E_g (eV)), and (g-i) ODCB-water partition coefficient (logP) of the whole C_{20} - C_{60} database as a function of the first moment pentagon signature P1 (left panel), the first moment hexagon signature H1 (middle panel), and the second moment hexagon signature H2 (right panel). The dashed lines correspond to a linear fit of the C_{60} isomer dataset.

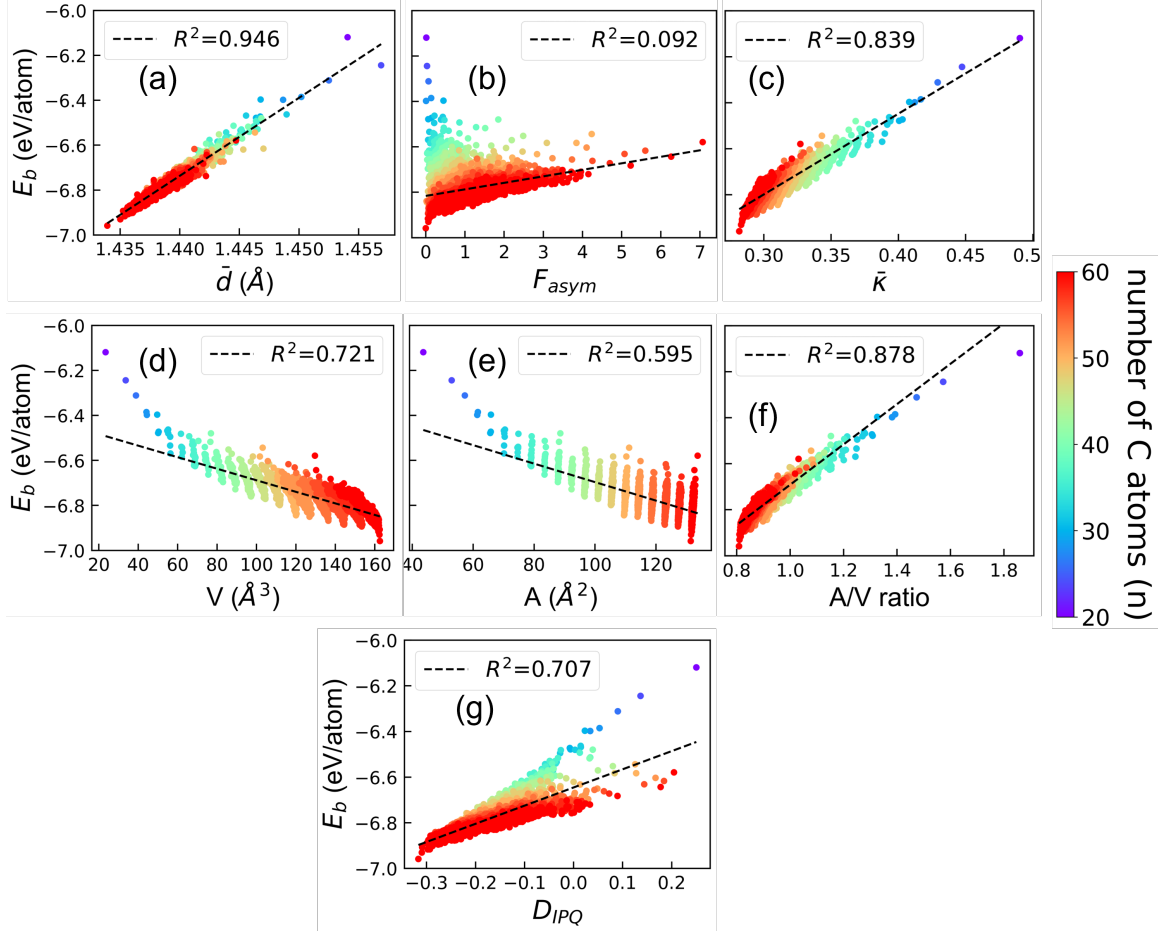


Figure S10: The binding energy values (E_b (eV/atom)) of the whole C_{20} - C_{60} database as a function of (a) average bond length $\bar{d}$ ($\text{\AA}$), (b) F_{asym} , (c) average curvature $\bar{\kappa}$, (d) volume V ($\text{\AA}^3$), (e) surface area A ($\text{\AA}^2$), (f) ratio of A/V , (g) D_{IPQ} . The dashed lines correspond to a linear fit of the whole C_{20} - C_{60} dataset. The color bar indicates varying fullerene sizes.

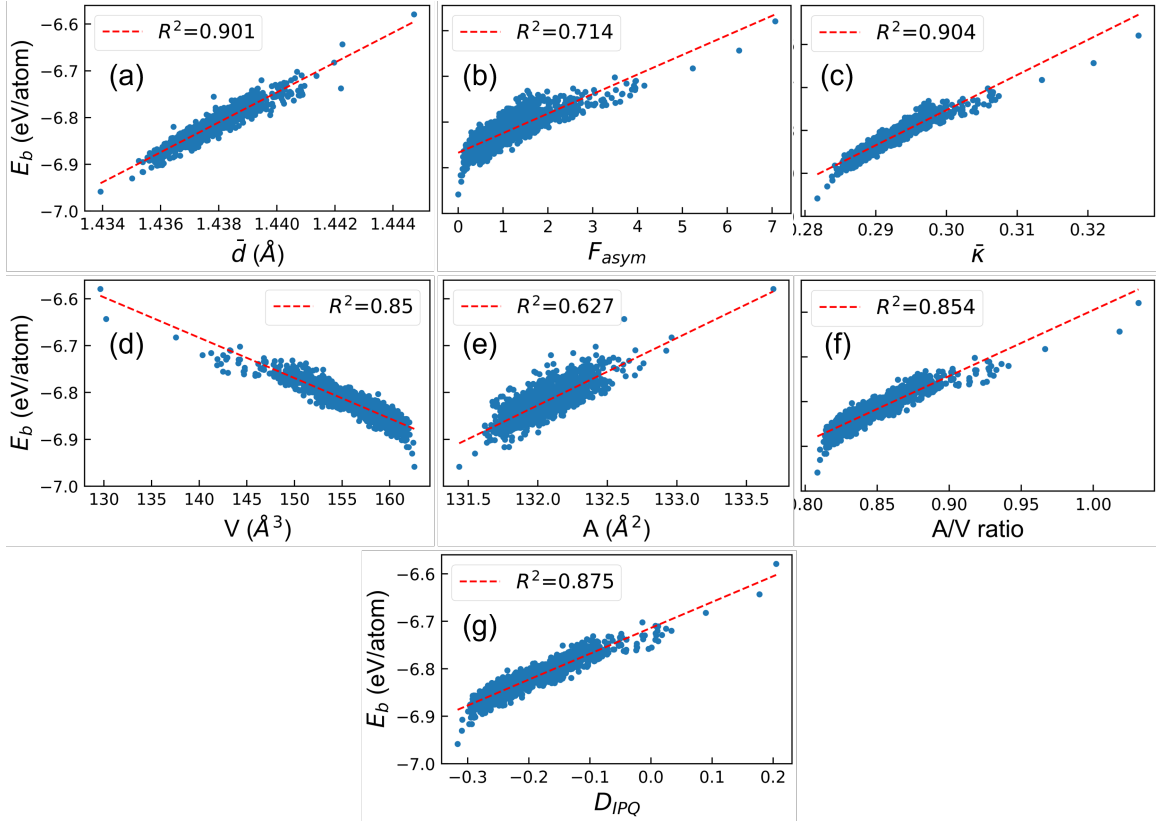


Figure S11: The binding energy values (E_b (eV/atom)) of the 1812 C_{60} database as a function of (a) average bond length $\bar{d}$ ($\text{\AA}$), (b) F_{asym} , (c) average curvature $\bar{\kappa}$, (d) volume V ($\text{\AA}^3$), (e) surface area A ($\text{\AA}^2$), (f) ratio of A/V , (g) D_{IPQ} . The dashed lines correspond to a linear fit of the C_{60} dataset.

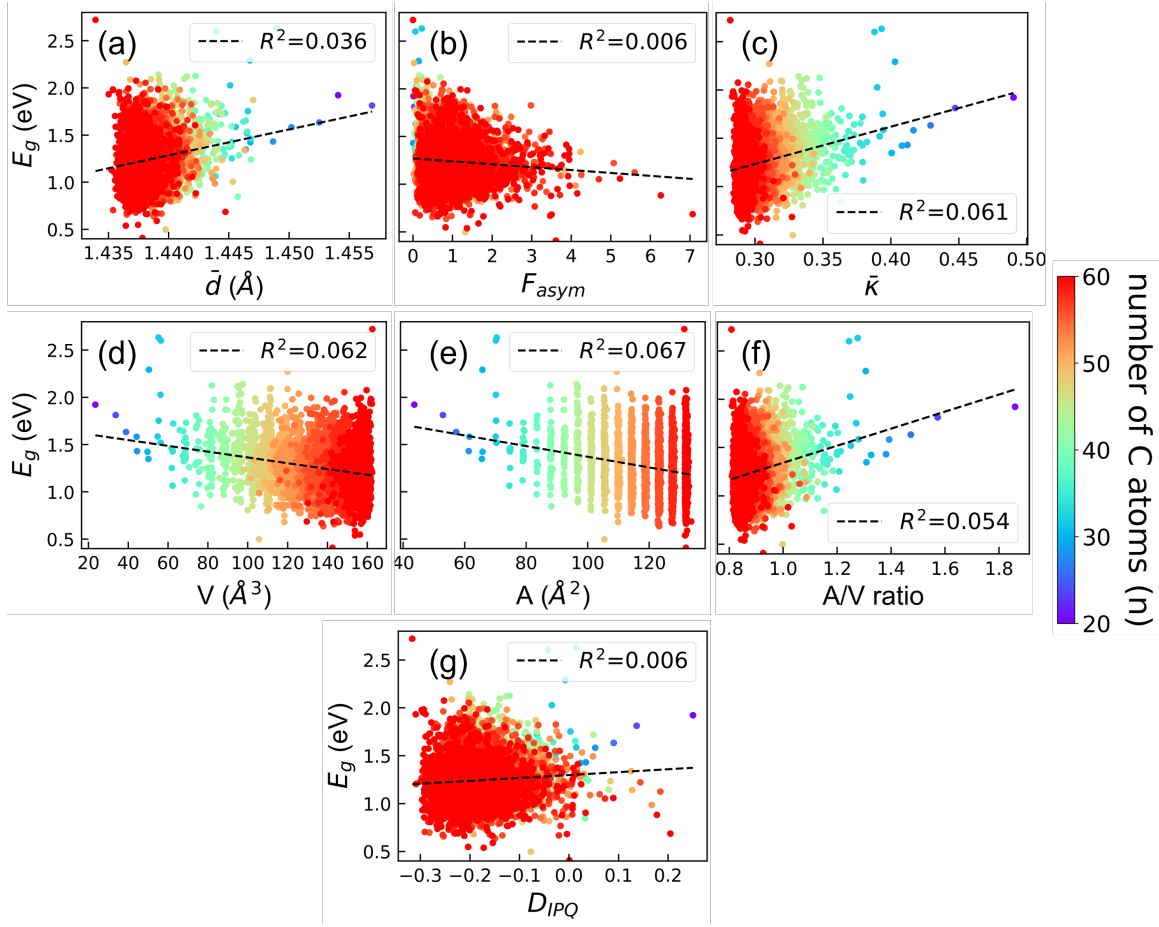


Figure S12: The HOMO-LUMO gap values (E_g (eV)) of the whole C_{20} - C_{60} database as a function of (a) average bond length $\bar{d}$ (Å), (b) F_{asym} , (c) average curvature $\bar{\kappa}$, (d) volume V (Å³), (e) surface area A (Å²), (f) ratio of A/V , (g) D_{IPQ} . The dashed lines correspond to a linear fit of the whole C_{20} - C_{60} dataset. The color bar indicates varying fullerene sizes.

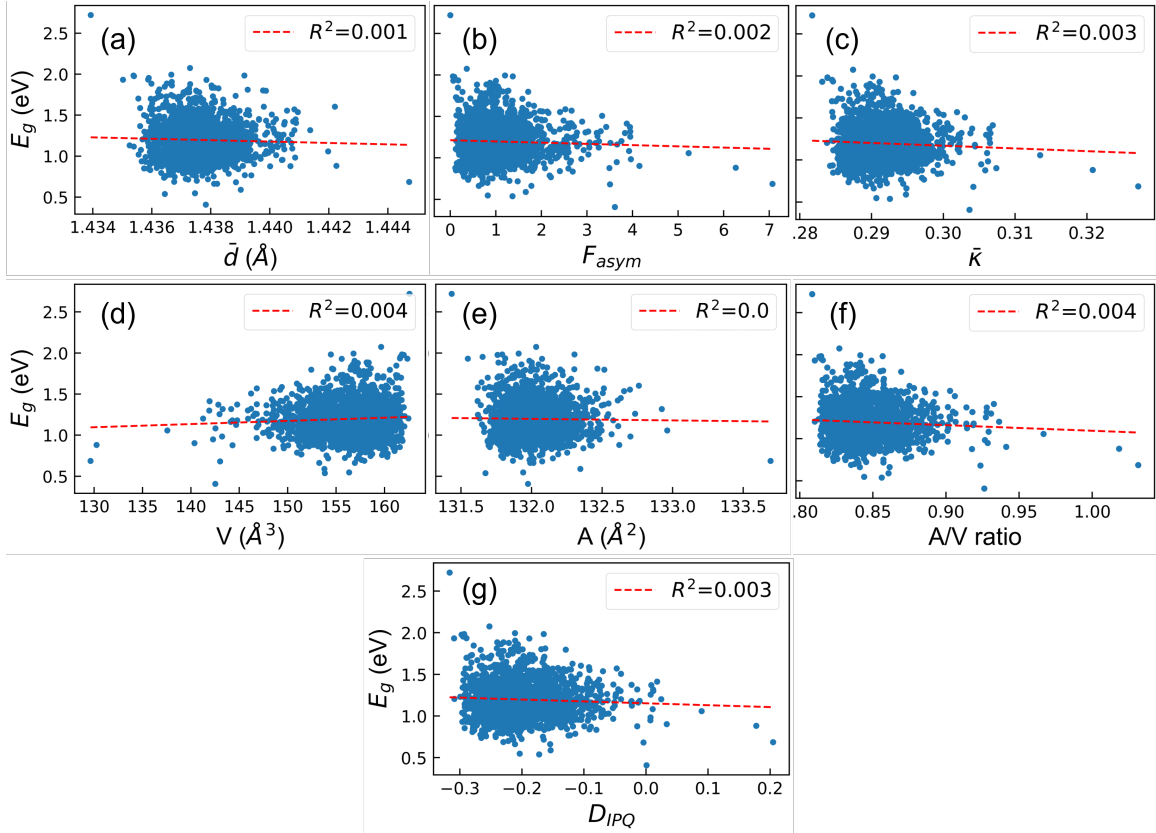


Figure S13: The HOMO-LUMO gap values (E_g (eV)) of the 1812 C_{60} database as a function of (a) average bond length $\bar{d}$ (Å), (b) F_{asym} , (c) average curvature $\bar{\kappa}$, (d) volume V (Å³), (e) surface area A (Å²), (f) ratio of A/V , (g) D_{IPQ} . The dashed lines correspond to a linear fit of the C_{60} dataset.

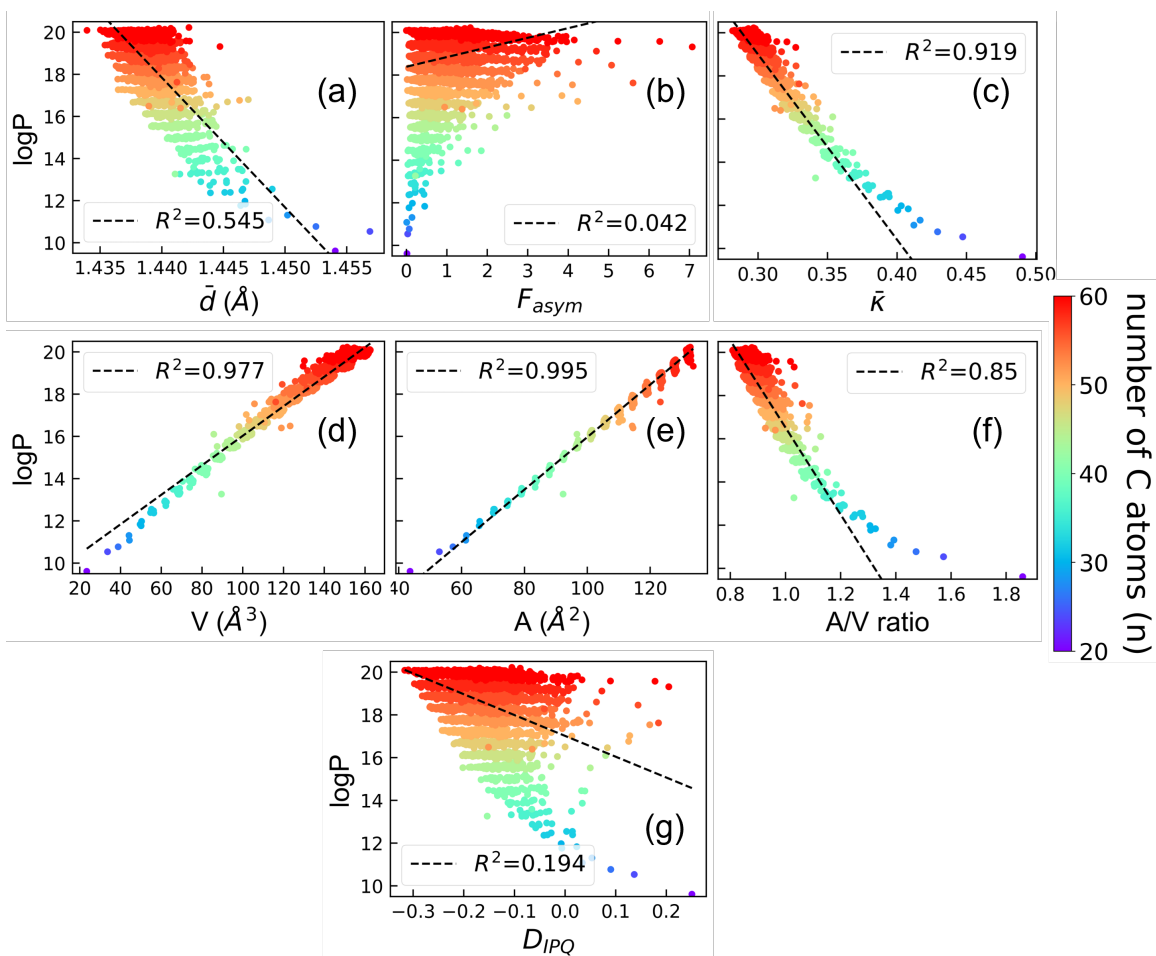


Figure S14: The ODCB-water partition coefficient values ($\log P$) of the whole C_{20} – C_{60} database as a function of (a) average bond length $\bar{d}$ (Å), (b) F_{asym} , (c) average curvature $\bar{\kappa}$, (d) volume V (Å³), (e) surface area A (Å²), (f) ratio of A/V , (g) D_{IPQ} . The dashed lines correspond to a linear fit of the whole C_{20} – C_{60} dataset.

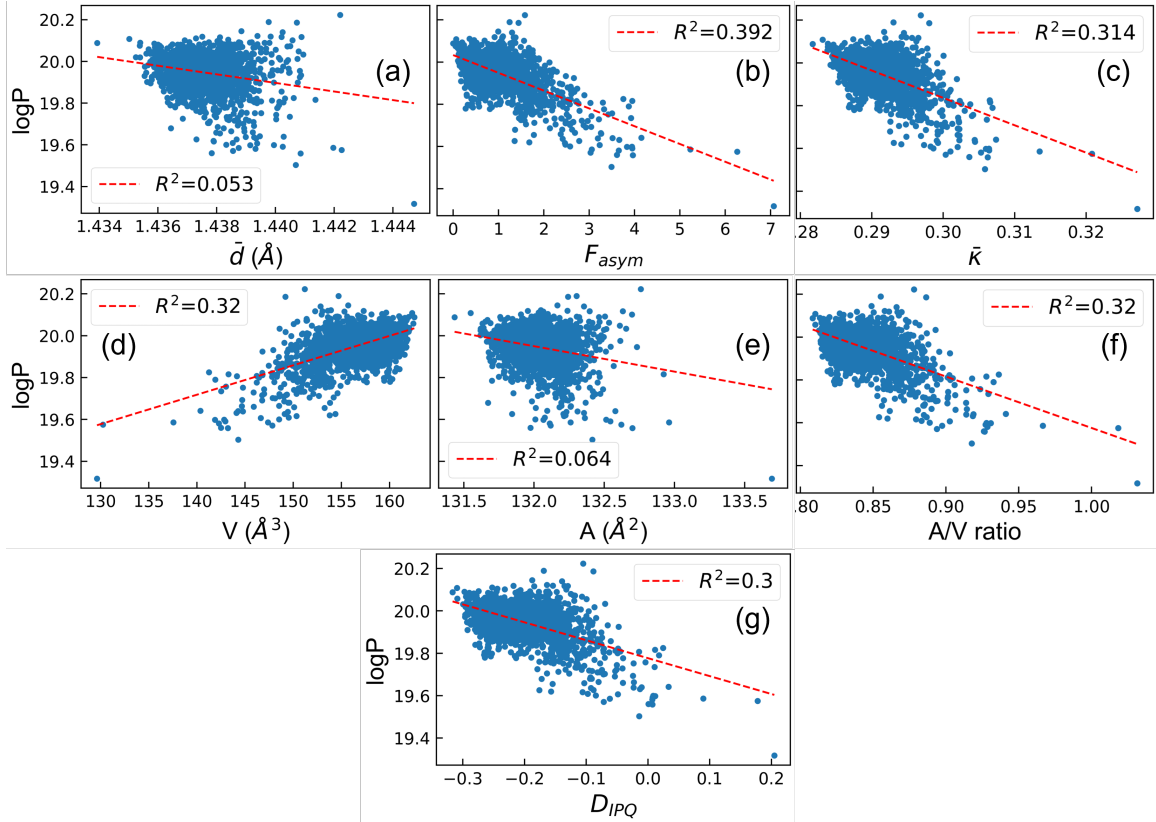


Figure S15: The ODCB-water partition coefficient values ($\log P$) of the 1812 C_{60} database as a function of (a) average bond length $\bar{d}$ ($\text{\AA}$), (b) F_{asym} , (c) average curvature $\bar{\kappa}$, (d) volume V ($\text{\AA}^3$), (e) surface area A ($\text{\AA}^2$), (f) ratio of A/V , (g) D_{IPQ} . The dashed lines correspond to a linear fit of the C_{60} dataset.

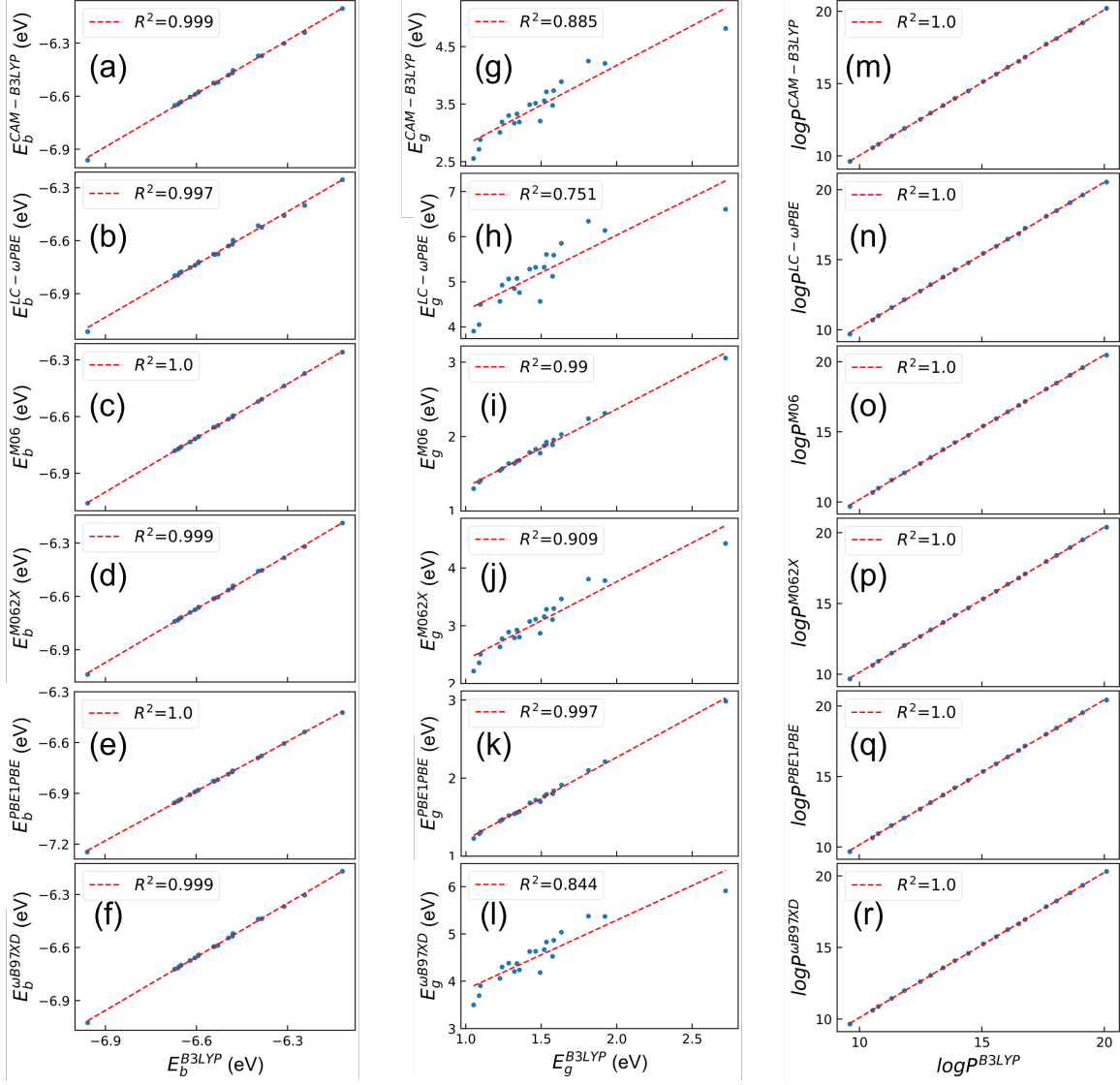


Figure S16: Correlation analysis between B3LYP and six other exchange-correlation functionals (CAM-B3LYP, LC- ω PBE, M06, M06-2X, PBE1PBE, and ω B97XD) in calculating (a-f) binding energy (E_b), (g-l) HOMO-LUMO gap (E_g), and (m-r) ODCB-water partition coefficient ($\log P$) for 20 isomer structures within C_{20} - C_{60} dataset, with each structure (C_n -#iso-1) representing a different cage size of C_n .

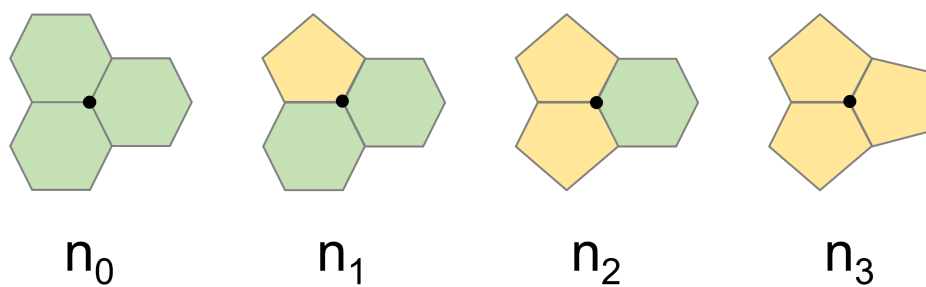


Figure S17: Schematic plot of 4 atom features.

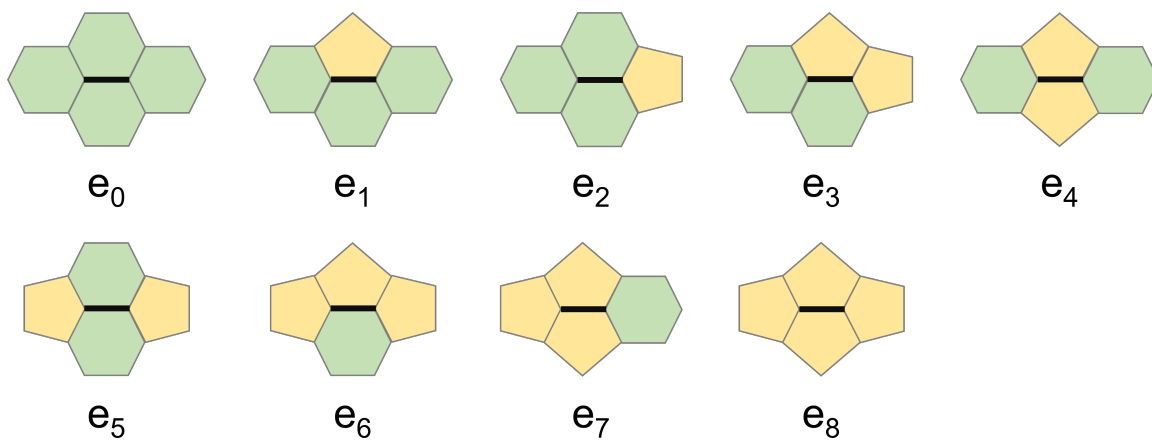


Figure S18: Schematic plot of 9 bond features.

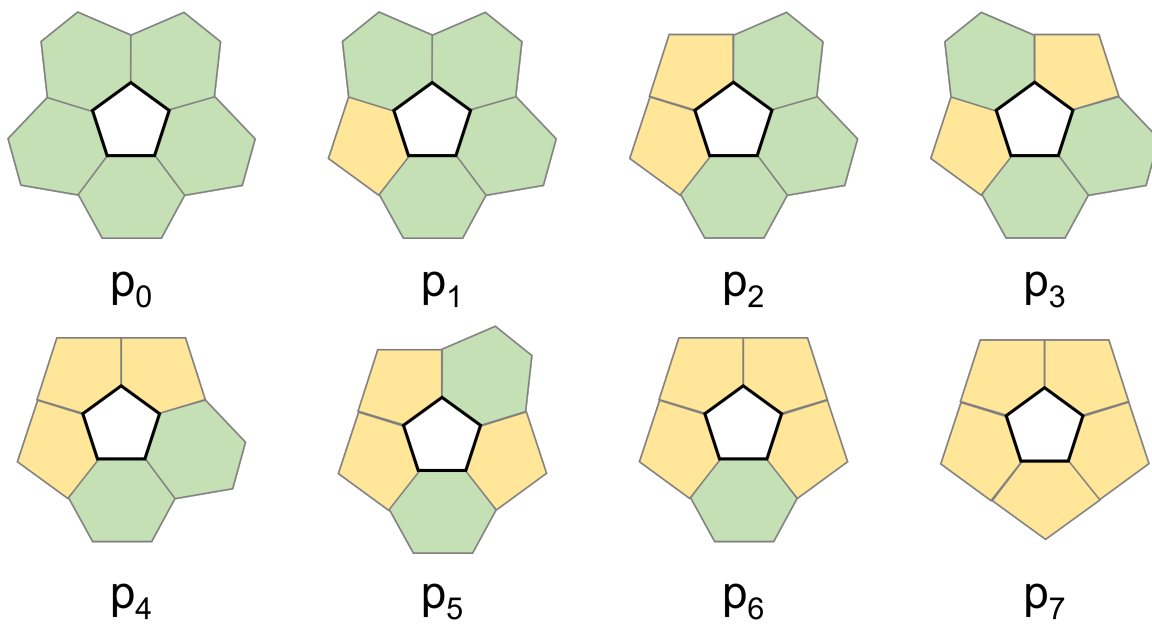


Figure S19: Schematic plot of 8 pentagon features.

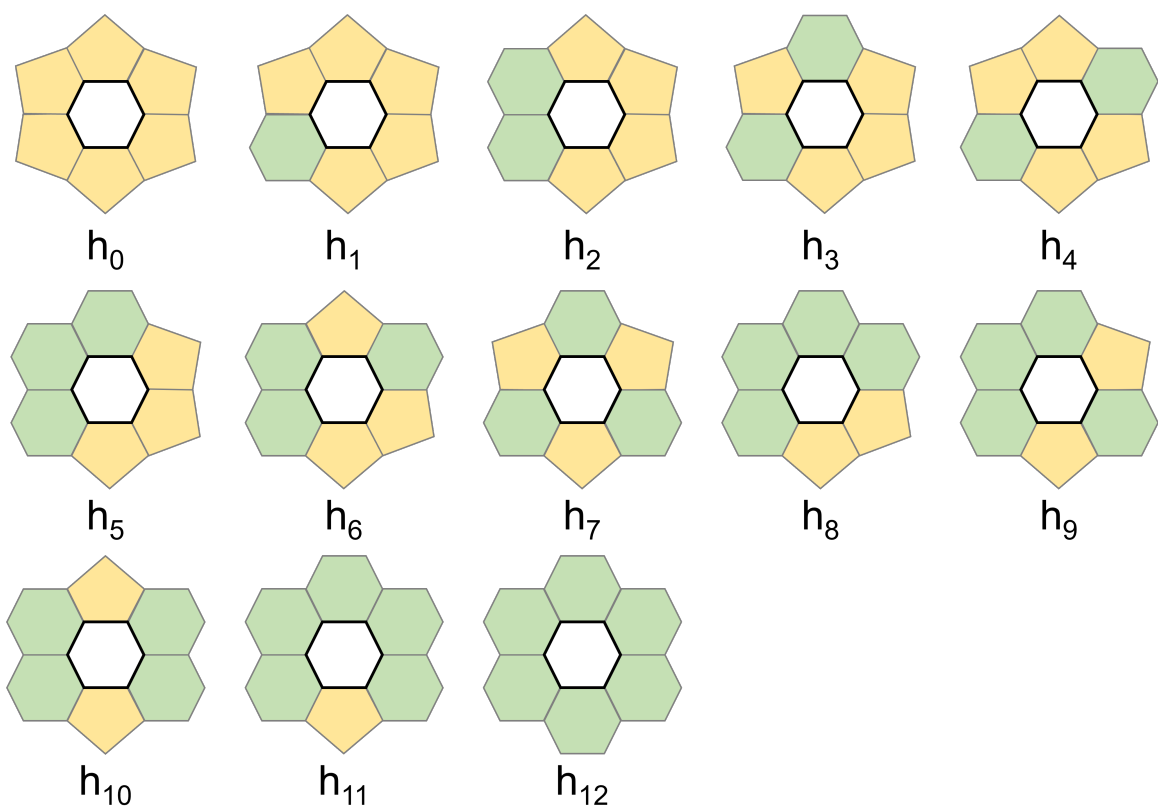


Figure S20: Schematic plot of 13 hexagon features.

Table S9: The values of 4 topological features, i.e., atom, bond, pentagon, and hexagon features, for three fullerene isomers (Cn), i.e., C₂₀-#1, C₆₀-#1 and C₆₀-#1812

	Cn	C ₂₀ -#1	C ₆₀ -#1	C ₆₀ -#1812
Atom features	n0	0	0	30
	n1	0	60	10
	n2	0	0	10
	n3	20	0	10
Bond features	e0	0	0	40
	e1	0	60	0
	e2	0	0	10
	e3	0	0	20
	e4	0	0	0
	e5	0	30	0
	e6	0	0	0
	e7	0	0	10
	e8	30	0	10
Pentagon features	p0	0	12	0
	p1	0	0	0
	p2	0	0	0
	p3	0	0	0
	p4	0	0	10
	p5	0	0	0
	p6	0	0	0
	p7	12	0	2
Hexagon features	h0	0	0	0
	h1	0	0	0
	h2	0	0	0
	h3	0	0	0
	h4	0	0	0
	h5	0	0	0
	h6	0	0	0
	h7	0	20	0
	h8	0	0	10
	h9	0	0	0
	h10	0	0	0
	h11	0	0	0
	h12	0	0	10

References

1. Sure, R., Hansen, A., Schwerdtfeger, P. & Grimme, S. Comprehensive theoretical study of all 1812 C₆₀ isomers. *Phys. Chem. Chem. Phys.* **19**, 14296–14305 (2017).