npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-024-01410-7

Mapping structure-property relationships in fullerene systems: a computational study from C_{20} to C_{60}

Bin Liu [1,2], Jirui Jin[1,2] & Mingjie Liu [1,2]

Fullerenes, as characteristic carbon nanomaterials, offer significant potential for diverse applications due to their structural diversity and tunable properties. Numerous isomers can exist for a specific fullerene size, yet a comprehensive understanding of their fundamental properties remains elusive. In this study, we construct an up-to-date computational database for C_{20}–C_{60} fullerenes, consisting of 5770 structures, and calculate 12 fundamental properties using DFT, including stability (binding energy), electronic properties (HOMO-LUMO gap), and solubility (partition coefficient logP). Our findings reveal that the HOMO-LUMO gap weakly correlates with both binding energy and logP, indicating that electronic properties can be tailored for specific uses without affecting stability or solubility. In addition, we introduce a set of topological features and geometric measures to investigate structure-property relationships. We apply atom, bond, and hexagon features to effectively predict the stability of C_{20}–C_{60} fullerenes, surpassing the conventional qualitative isolated pentagon rule, and demonstrating their robust transferability to larger-size fullerenes beyond C_{60}. Our work offers guidance for optimizing fullerenes as electron acceptors in organic solar cells and lays a foundational understanding of their functionalization and applications in energy conversion and nanomaterial sciences.

Carbon-based nanomaterials, including fullerenes, nanotubes, and graphene, are widely utilized in various fields for their unique properties that arise from their aromatic carbon frameworks[1–5]. Among them, fullerenes are polyhedral cage molecules with the general molecular formula $C_{20+2n}(n \geq 0, n \neq 1)$ composed of sp^2 hybridized carbons arranged in pentagon and hexagon rings. The variation in the arrangement of these rings across the fullerene cages leads to numerous isomeric structures for each fullerene size. For example, C_{60} fullerene adopts 1812 non-isomorphic structures with different distributions of the 12 pentagon and 20 hexagon rings on a spherical surface. The number of possible isomers for fullerene cages with $N = 20 + 2n$ carbon atoms increases at a rate of $O(N^9)$[6], creating an extensive family of fullerenes with a vast array of isomers. The unique zero-dimensional spherical cage structures of fullerenes contribute to their remarkable physical and chemical characteristics, such as high thermal and electrical conductivity, extraordinary tensile strength, and efficient electron donation and acceptance abilities[7]. Owing to their structural diversity that offers tunable electronic properties and reactivities, fullerenes have been extensively investigated as components in optoelectronics[8–11], solar cells[12–14], gas storage and separation[15,16], biology and medicine[17–19]. Although fullerenes have significant untapped potential in various applications, fully harnessing their capabilities—especially in customizing electronic properties and functionalization for practical uses—requires a thorough understanding of the structure-property relationships across the entire spectrum of fullerene varieties.

The investigation of fundamental properties across different fullerene isomers has been the focus of numerous earlier studies[20–28]. For example, Rebecca et al. investigated the relative energy distribution among the 1812 isomers of C_{60} fullerene with the DFT method and correlated the isomeric stability with a variety of topological indices, electronic, and geometric properties. Aside from the common isolated pentagon rule (IPR), they proposed that a small pentagon signature P_1, a large volume, and a more spherical cage can lead to a relatively stable isomer[23]. Zhao et al. explored the origin and characteristics of isomeric stability in four fullerene systems, C_{44}, C_{48}, C_{52}, and C_{60}, using DFT simulations. Through energy decomposition analysis, they found that electrostatic potential is the primary determinant of isomeric stability, surpassing other factors such as steric and quantum effects. They further highlighted the importance of spatial delocalization of the electron density on the stabilization of fullerene isomers[29]. Chan and

[1]Department of Chemistry, University of Florida, Gainesville, FL 32611, USA. [2]Quantum Theory Project, University of Florida, Gainesville, FL 32611, USA.
e-mail: mingjieliu@ufl.edu

Table 1 | The summary of 12 fundamental properties for the 5770 C_{20}–C_{60} database calculated at B3LYP level

Fundamental properties	Unit
Binding energy (E_b)	eV/atom
HOMO level	eV
LUMO level	eV
HOMO-LUMO gap (E_g)	eV
Ionization potential (IP)	eV
Electron affinity (EA)	eV
IP-EA gap (E_{fund})	eV
Dipole moment (μ)	D
Triplet-singlet energy difference (ΔE_{T-S})	eV
Solvation free energy in water ($\Delta G_{sol}(water)$)	kJ/mol
Solvation free energy in ODCB ($\Delta G_{sol}(ODCB)$)	kJ/mol
ODCB–water partition coefficient (logP)	–

Karton studied the size dependence of a variety of electronic properties in small fullerenes from C_{20} to C_{50} with 812 isomeric structures, including the energy difference between the lowest-energy singlet and triplet states (i.e., S_0 and T_1, or T_0 and S_1), and the related quantities of ionization potential (IP) and electron affinity (EA). They found a linear correlation between triplet-singlet energy difference and IP-EA difference and further surveyed larger fullerenes in search of candidates with superior charge-transfer properties[30]. Based on our comprehensive literature review, we observed that the previous research into the structure-property relationships of fullerenes has been grounded in selected datasets that represent an incomplete array of fullerene isomers, with the bulk of these studies primarily focusing on stability analysis. To the best of our knowledge, there remains a gap in the literature for a comprehensive and accurate examination of the fundamental properties of all fullerene isomers from C_{20} to C_{60}.

In this study, we have compiled the most extensive computational dataset of C_{20}–C_{60} fullerenes to date, incorporating a total of 5770 structures. Following our benchmark study, we calculated 12 fundamental properties with DFT-level accuracy, e.g., fullerene binding energy (E_b), HOMO-LUMO gap (E_g), and partition coefficient of solvation free energies in 1,2-dichlorobenzene (ODCB) and water phases as listed in Table 1, and evaluated their Pearson correlation coefficients. The weak correlations between E_g and both E_b and logP suggest that favorable electronic properties can be independently adjusted for particular applications without compromising stability and solubility. To delve deeper into the structural features influencing these properties, we introduced various topological indices and geometric measures, going beyond the IPR. We reveal that compared to pentagon-related features, atom, bond, and hexagon features demonstrate significantly superior effectiveness in capturing the intricate local structural environments of the carbon atoms on the spherical cage. The linear models fitted by those features enable the predictions of the fullerene stability across various sizes and show robust transferability to larger-size fullerenes beyond C_{60}. We further offer a practical guide for researchers interested in screening fullerenes for optimal electron acceptor candidates in organic solar cells. Given the prevalence of data-driven models developed in recent decades, we emphasize that our comprehensive datasets are crucial for model training and new materials discovery in the fullerene field. Our study provides a deep understanding of the structure-property relationships across fullerene isomers, setting a foundation for future advancements in fullerene functionalization and applications in energy conversion and nanosciences.

Results and discussion

Stability analysis

We calculated the binding energy per carbon atom using the atomic carbon in the triplet state as the reference[31], which serves as one indicator of the thermodynamic stability of 5770 fullerene structures. However, other parameters to assess thermodynamic stability, such as addition energy[32] and activation barrier[33], which require more computationally intensive calculations, were not included in this study.

$$E_b(C_n) = \frac{1}{n}(E_{C_n} - nE_C) \quad (1)$$

E_{C_n} is the total energy of fullerene C_n with n carbon atoms, whereas E_C is the energy of an isolated carbon atom at triplet state. The binding energy distribution indicates a trend of decreasing binding energies as the fullerene cage size increases, as depicted in Fig. 1a. Within the dataset, the renowned C_{60}–Ih buckminsterfullerene (C_{60}–#1), which possesses icosahedral symmetry and is the sole isomer conforming to the IPR, exhibits the lowest E_b of −6.96 eV/atom, comparable to the experimental value of −7.04 eV/atom[34]. Conversely, the C_{20}–#1 fullerene, comprised solely of 12 pentagons devoid of hexagons, exhibits the highest E_b of −6.12 eV/atom. The statistical analysis of the E_b values is detailed in Table S1, together with the molecular geometries of fullerene isomers shown in Fig. S1.

To assess the accuracy of the DFT-calculated energies, we benchmarked the relative energies of ten C_{60} isomers compared to C_{60}–#1 with the relative energies from DLPNO-CCSD(T)/CBS* reference results adopted from Rebecca et al.'s work[23]. The geometries of these ten isomers are shown in Fig. S2. The DLPNO-CCSD(T)/CBS* relative energies are expected to have an error of ±0.5 kcal/mol, which can be used for benchmark[35,36]. The comparison reveals that the outcomes obtained using B3LYP-D3/6-311G* exhibit a consistent ranking of relative energies for the ten fullerene isomers with small deviations when benchmarked against the DLPNO-CCSD(T)/CBS* results (Table S2). To ensure statistical significance, we further selected one isomer from each fullerene size from C_{20} to C_{58} for E_b calculations at the B3LYP-D3/6-311G* level to benchmark against DLPNO-CCSD(T)/CBS*. As shown in Table S3, the relative binding energies (ΔE_b) with respect to C_{60}–#1 obtained using B3LYP/6-311G* are consistent with the DLPNO-CCSD(T)/CBS* results across a diverse set of fullerenes.

The histogram illustrating the binding energy distribution among the 5770 fullerene structures reveals that 80% are within the energy range of −6.85 to −6.72 eV/atom, with an average binding energy of −6.79 ± 0.06 eV/atom (Fig. 2a). This indicates that most of the fullerene structures are much more stable than C_{20}. Prior research has shown that C_{20} can be synthesized in the gas phase[37], indicating the possibility of synthesizing other more stable fullerene structures. In addition, other smaller fullerenes like C_{36} have been synthesized experimentally using the arc-discharge method[38]. Functionalization, both endohedral and exohedral, has been shown to enhance the stability of non-IPR fullerene isomers, exemplified by compounds such as $C_{50}Cl_{10}$[39], $C_{60}Cl_{12}$, and $C_{60}Cl_8$[40]. We anticipate that our comprehensive computed binding energy will provide a reliable benchmark for future theoretical and experimental research into fullerene synthesis and functionalization.

Electronic property analysis

Frontier molecular orbitals, i.e., HOMO and LUMO levels, along with HOMO-LUMO gap (E_g), are critical properties in evaluating the electronic and optical properties, electric conductivity, charge transport, and chemical reactivity of materials. Specifically, understanding the distribution of these fundamental properties across various sizes of fullerenes is essential for designing and optimizing fullerenes and their derivatives with specific requirements, such as improved conductivity, enhanced light absorption, or more favorable reactivity. In contrast to the distribution of E_b values, the distribution of E_g exhibits a lower dependence on the size of the fullerene cage, as shown in Fig. 1b. The histogram plot in Fig. 2b reveals that approximately 80% of E_g values are within the range of 0.97–1.54 eV, with the maximum gap being 2.72 eV (C_{60}–#1) and minimum at 0.41 eV (C_{60}–#1748), calculated at the B3LYP level (also see Table S1 and Fig. S1). The distributions of HOMO and LUMO also exhibit weak dependence on the fullerene size, and the details can be found in Fig. S3.

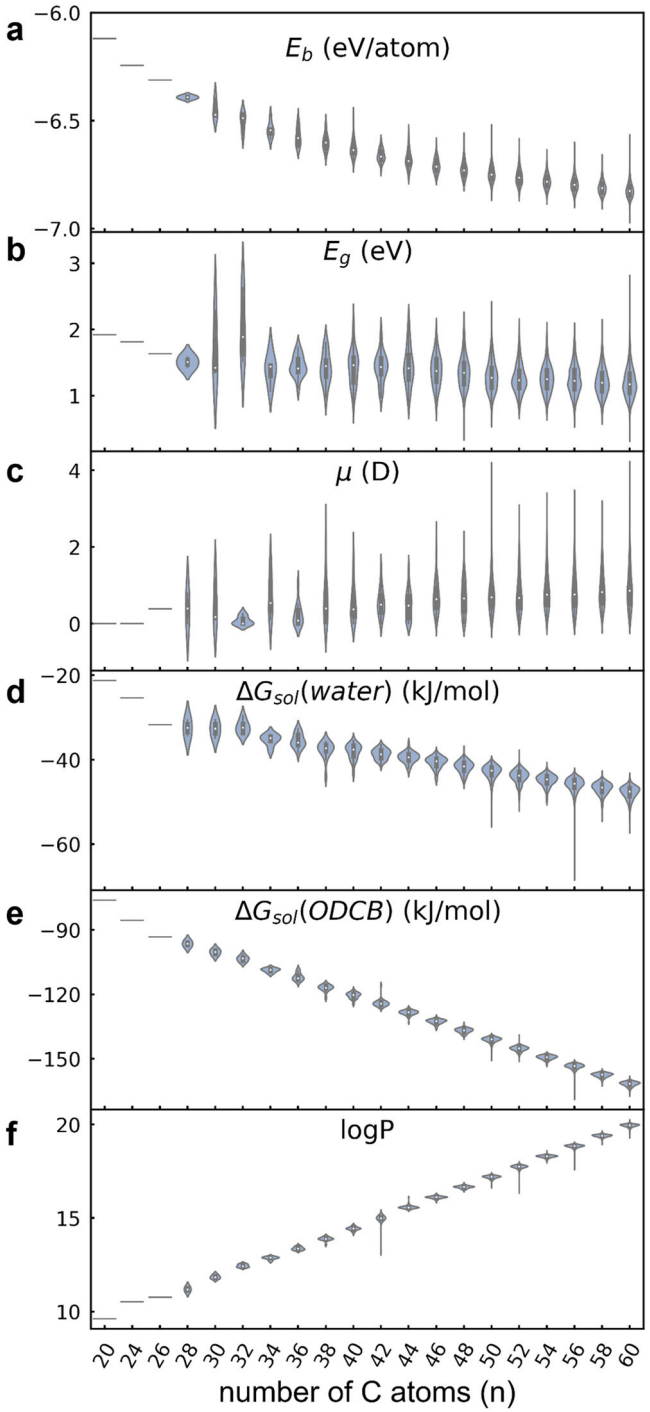

Fig. 1 | Size-dependence analysis on 6 fundamental properties. Violin plots for the distributions of **a** binding energy E_b (eV/atom), **b** HOMO-LUMO gap E_g (eV), **c** dipole moment μ (D), **d**, **e** solvation free energy in water $\Delta G_{sol}(water)$ and 1,2-dichlorobenzene (ODCB) $\Delta G_{sol}(ODCB)$ (kJ/mol), **f** ODCB–water partition coefficient logP from fullerene C_{20} to C_{60}.

Currently, fullerene C_{60}, C_{70}, C_{84} and their derivatives have been widely explored as the acceptor to pair with conjugated polymer donors in organic solar cells (OSCs)[12,41–43]. To offer a practical guide for researchers in the field of OSCs, we applied Scharber's model[44–46] to assess the photovoltaic performance of the fullerenes. The power conversion efficiency (PCE) of an OSC device under sunlight irradiation is determined by open-circuit voltage (V_{OC}), short-circuit current density (J_{SC}), fill factor (FF), and the power of incident light P_{in} with:

$$PCE = \frac{P_{out}}{P_{in}} = \frac{V_{OC} J_{SC} FF}{P_{in}} \times 100\% \quad (2)$$

P_{in} is calculated by integrating the AM1.5 spectra[47] across the wavelength, resulting in an approximate value of 1000 W/m^2. V_{OC} can be determined by the HOMO level of polymer donor (E^D_{HOMO}) and the LUMO level of fullerene acceptor (E^A_{LUMO}) as follows:

$$V_{OC} = \frac{1}{e}|E^D_{HOMO} - E^A_{LUMO}| - 0.3 \quad (3)$$

J_{SC} is assumed to be the current from absorbing all the incident photons above the E_g of the conjugated polymer donor as follows:

$$J_{SC} = EQE \int_{E_g}^{\infty} e\Phi_{ph}(E)dE \quad (4)$$

in which the external quantum efficiency EQE is set to 0.65, and Φ_{ph} represents the incident solar photon flux density as a function of energy E. In addition, FF in eq(2) is set to 0.65. Furthermore, in fullerene-based OSCs, energy offsets exceeding 0.3 eV between the HOMO levels of the polymer donor and fullerene (ΔE_{HOMO}); as well as between the LUMO levels of the polymer donor and fullerene (ΔE_{LUMO}), are essential to provide the necessary driving force for efficient exciton dissociation[42]. Scharber's model has been extensively applied in high-throughput screening and the design of OSC device[48], such as the Harvard Clean Energy Project database introduced by Hachmann et al.[46].

Here we employ the widely used polymer P3HT as an example of donor material. We utilized an oligomer comprising 8 thiophene rings to depict the polymer, while also substituting all alkyl side chains with hydrogen atoms as shown in Fig. 3b, which exhibits a HOMO level of −4.85 eV and a LUMO level of −2.48 eV at B3LYP level from our calculations. Accordingly, the ideal fullerene acceptor should possess a HOMO level of approximately −5.15 eV and a LUMO level of −2.78 eV (inset of Fig. 3a). The estimated PCE of C_{60}–Ih (C_{60}–#1) based OSC is about 8.1%, which is consistent to the experimental value of ~5%[49,50]. By applying the aforementioned requirements as screening conditions to the fullerene database, 2771 out of 5770 fullerene isomers could be potential candidates to pair with P3HT in OSC devices (see Fig. 3a). The highest PCE values can reach 9.6%, achieved by C_{20}–#1, as well as the next three top-performing fullerene structures showed in Fig. 3b.

Dipole moment stands as another critical property for OSC materials, which affects intermolecular interactions and the morphology of the organic material film. Previous studies have highlighted the effectiveness of strategically selecting and designing materials with appropriate dipole moments to enhance charge separation, promote efficient exciton dissociation, optimize energy level alignment at interfaces, and consequently improve the overall efficiency and stability of OSCs[51–54]. The dipole moment distribution of fullerenes from C_{20} to C_{60} also shows weak dependence on the size of the fullerene cage (Fig. 1c). The majority of fullerenes display a permanent dipole moment that is non-negligible. For example, among 5770 fullerene structures, only around 200 exhibit a zero dipole moment, as shown in Fig. 2c. Notably, fullerene C_{60}–#1646 exhibits the largest dipole moment of 3.98 D among all the fullerenes (see Table S1).

In addition to E_g, other fundamental quantities related to electronic properties have also been computed, such as the fundamental gap E_{fund} between IP and EA, and the energy difference between triplet and singlet states ΔE_{T-S}, which are also crucial for the stability and catalytic activity of materials. As shown in Fig. S4, E_g values show strong correlations with both E_{fund} and ΔE_{T-S} across the C_{20}–C_{60} database, with R^2 values exceeding 0.9,

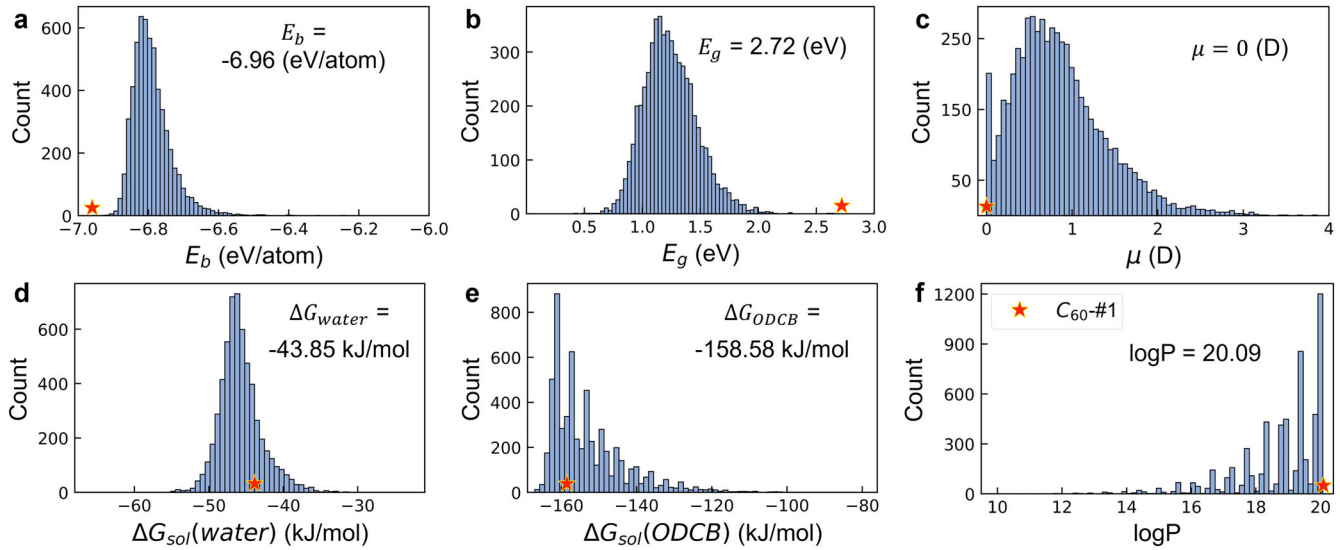

Fig. 2 | Histogram distributions of 6 fundamental properties. Histogram plots for the distributions of **a** binding energy E_b (eV/atom), **b** HOMO-LUMO gap E_g (eV), **c** dipole moment μ (D), **d**, **e** solvation free energy in water $\Delta G_{sol}(water)$ and 1,2-dichlorobenzene (ODCB) $\Delta G_{sol}(ODCB)$ (kJ/mol), **f** ODCB–water partition coefficient logP from fullerene C_{20} to C_{60}. The star marks highlight the values of C_{60}–#1.

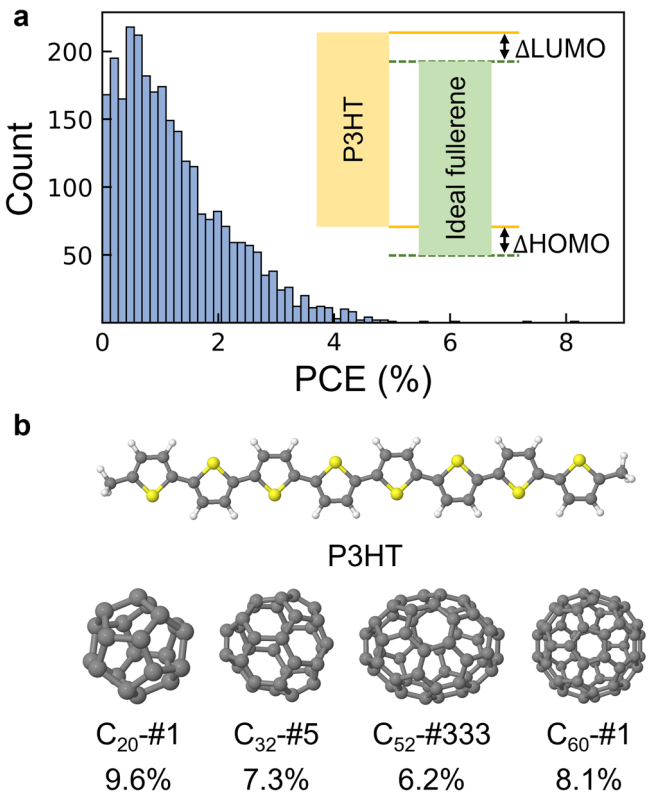

Fig. 3 | Estimated efficiency of fullerene-based organic solar cells. a The power conversion efficiency (PCE) of fullerene-based OSCs with P3HT as donor estimated by Scharber's model (inset: Energy level diagram for ideal fullerene acceptors aligned with P3HT polymer donor in OSC devices. The HOMO level offsets (ΔE_{HOMO}) and LUMO level offsets (ΔE_{LUMO}) between P3HT and fullerenes are ≥0.3 eV). **b** Chemical structures of P3HT oligomer and four fullerenes with the highest PCE values predicted by Scharber's model (plotted with Jmol).

whereas the linear correlation between the fundamental gap and triplet-singlet energy difference exhibits cage size dependence with a lower R^2 of 0.774. Indeed, Table S4 reveals that, for fullerene cages of the same size, R^2 values for the E_g, E_{fund}, and ΔE_{T-S} all exceed 0.94, with the exception of C_{36}, which are consistent with the previous studies[30].

Besides its significance in photovoltaic applications, E_g is also connected to other essential properties in electrochemistry. For instance, E_g is linked to the redox potential, a key factor in processes such as cyclic voltammetry (CV), energy storage in batteries and fuel cells, and electrocatalysis. Previous studies have demonstrated that, across a series of five fullerenes, the experimentally determined redox potential strongly correlates with the calculated E_g[55]. We are confident that our extensive dataset of fullerenes, with computed HOMO, LUMO, and E_g, can assist in the selection of potential fullerene candidates for solar cells and electrochemical applications.

Solubility and ODCB–water partition coefficient analysis

Solubility is an essential factor for the practical utilization of fullerenes. It is widely recognized that fullerenes are insoluble in water but exhibit high solubility in organic solvents[56,57]. For example, the solubility of C_{60} is 24 g/L in 1,2-dichlorobenzene (ODCB) solution, compared to 1.3×10^{-11} g/L in water[57]. In this work, we calculated the Gibbs free energy of solvation (ΔG_{sol}) in ODCB and water for our fullerene dataset. The ΔG_{sol} is computed as the electronic energy difference between the fullerene molecule in solvent phase $E(solvent)$ and gas phase $E(gas)$ as given in the equation with other terms in free energies ignored:[58]

$$\Delta G_{sol} = E(solvent) - E(gas) \quad (5)$$

We observed a linear correlation between both $\Delta G_{sol}(ODCB)$ and $\Delta G_{sol}(water)$ with the size of fullerene cage; larger sizes correspond to lower solvation energies (Fig. 1d, e). Previous studies have also found that the solvation free energies of carbon allotropes, including C_{60} and carbon nanotubes, exhibit a negative linear correlation with their surface area[59]. The correlation between solvation free energy and geometric measures will be addressed in a later section.

The histogram depicting the distribution of $\Delta G_{sol}(water)$ in Fig. 2d indicates that around 80% of $\Delta G_{sol}(water)$ values are within the range of −48.81 to −42.10 kJ/mol with an average value of −45.74 ± 2.88 kJ/mol (also see Table S1). The maximum $\Delta G_{sol}(water)$, observed for C_{20}–#1 is −21.32 kJ/mol, while C_{56}–#3 exhibits the minimum at −67.74 kJ/mol. Particularly, the $\Delta G_{sol}(water)$ value for C_{60}–#1 molecule is −43.85 kJ/mol.

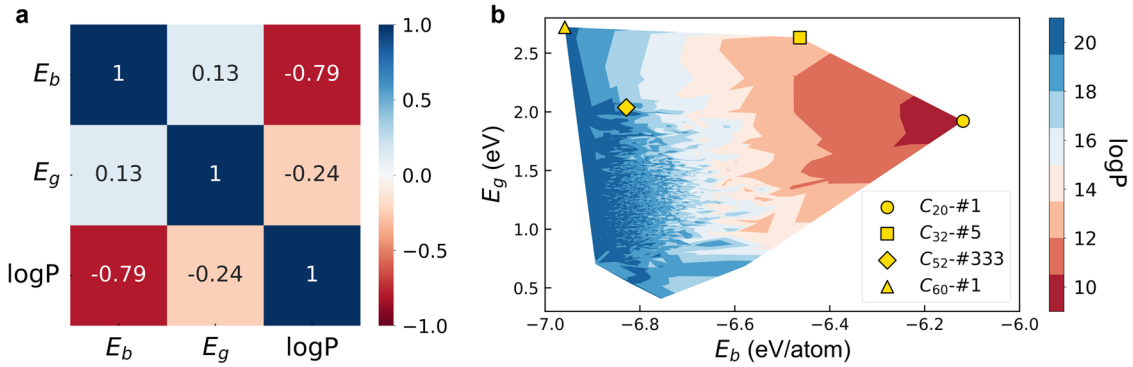

Fig. 4 | Correlation analysis among stability, electronic property, and solubility of fullerenes. a Pearson correlation coefficients and **b** contour plot of binding energy E_b (eV/atom), HOMO-LUMO gap E_g (eV), and logP. The yellow dots highlight the values of four fullerenes with the highest power conversion efficiency (PCE) values predicted by Scharber's model. (The chemical structures are depicted in Fig. 3b).

Previous research using various computational methodologies reported values are −55.27 kJ/mol by Varanasi et al.[60], −54.1 kJ/mol by Garde et al.[61], −50.9 kJ/mol by Kevin et al.[59], −36.10 kJ/mol by Muthukrishnan et al.[62], −18.4 kJ/mol by Graziano[63], and −2.9 kJ/mol by Evgeny et al.[64] It is important to note that these values exhibit significant variance, with most diverging considerably from the experimental value of −17.4 kJ/mol[64], underscoring the challenges in accurately predicting the $\Delta G_{sol}(water)$ values for fullerenes. In contrast, the histogram illustrating the distribution of $\Delta G_{sol}(ODCB)$ in Fig. 2e shows a skewed pattern, suggesting a pronounced preference of fullerene molecules for organic solvents. Approximately 80% of the value falls within the range of −162.08 to −139.90 kJ/mol (Fig. 2e). Similar to $\Delta G_{sol}(water)$, C_{20}–#1 and C_{56}–#3 exhibit the highest and lowest $\Delta G_{sol}(ODCB)$ values of −76.21 and −168.41 kJ/mol, respectively (see Table S1). Meanwhile, C_{60}–#1 shows a relatively low $\Delta G_{sol}(ODCB)$ value of −158.58 kJ/mol.

To indicate the partitioning of fullerenes between ODCB and water phases, we calculated the ODCB–water partition coefficient logP as given in equation:[65,66]

$$\log P = -\frac{\Delta G_{sol}(ODCB) - \Delta G_{sol}(water)}{2.303RT} \quad (6)$$

in which R is the gas constant (8.31 J/(mol · K)) and T is the room temperature (298.15 K). The partition coefficient serves as an indicator for predicting the lipophilicity of molecules. A higher logP value suggests a stronger propensity for dissolution in non-polar organic solvents as opposed to the polar aqueous phase. For fullerenes, logP demonstrates a significant correlation with cage size, as illustrated in Fig. 1f. Fullerene molecules with favorable solubility in ODCB typically exhibit logP values significantly greater than 0. Among the 5770 fullerene isomers, 80% of logP values fall in the range of 17.1–19.99 with an average value of 18.8 ± 1.32 (Fig. 2f). The maximum and minimum logP values are 20.22 for C_{60}–#1790 and 9.61 for C_{20}–#1, respectively (see Table S1). The logP value for C_{60}–#1 stands at 20.09, with $\Delta G_{sol}(water)$ and $\Delta G_{sol}(ODCB)$ calculated as −43.85 and −158.58 kJ/mol, respectively, signifying a pronounced preference of C_{60}–#1 molecule for the ODCB solvent over water. This finding is consistent with the previous studies. For example, Evgeny et al. reported the $\Delta G_{sol}(water)$ and $\Delta G_{sol}(ODCB)$ as −2.9 and −123.9 kJ/mol, respectively, calculated by effective Hamiltonian methods, which gives the logP value of 21.19. They also obtained the experimental solvation free energies of −17.4 and −127.6 kJ/mol for C_{60}–#1 in water and ODCB, which results in a logP value of 19.29[64]. Our findings suggest that while the absolute ΔG_{sol} values may vary from experimental data, the logP value–derived from the relative difference between $\Delta G_{sol}(water)$ and $\Delta G_{sol}(ODCB)$–consistently reflects the lipophilicity of molecules.

For the practical application of fullerenes in energy conversion, storage, electrochemistry, and nanoelectronics, it is essential to concurrently evaluate multiple key properties. Here we examine the Pearson correlation coefficients (r) among 12 fundamental properties for C_{20}–C_{60} fullerenes (see Fig. S5). Particularly, the correlations among stability, electronic properties, and solubility were assessed by calculating the r values among E_b, E_g, and logP. As shown in Fig. 4a, E_b is inversely related to logP with $r = -0.79$. This suggests that the greater stability, indicated by a lower binding energy, is associated with enhanced solubility in the ODCB phase. In comparison, the E_g value exhibits a weak correlation with both binding energy and logP, with $r =$ 0.13 and −0.24, respectively. These findings imply that it is possible to attain high stability, desirable solubility, and an optimal E_g concurrently for targeted applications. The density plot in Fig. 4b encapsulates the correlation between these properties and facilitates the identification of potential candidates for customizing properties to suit specific applications. For instance, when considering four fullerenes with the highest PCE values predicted by Scharber's model (see Fig. 3b), C_{52}–#333 and C_{60}–#60 emerge as more suitable fullerene acceptors compared to C_{20}–#1 and C_{32}–#5 owing to their superior stability and solubility.

Correlations of topological features and indices with fundamental properties

So far, we have constructed a comprehensive dataset of C_{20}–C_{60} fullerenes with 5770 structures and provided 12 DFT-level fundamental properties, including E_b, E_g, and logP. Previous studies have shown that the IPR can be employed to identify a set of stable isomers among all possible isomers for a given fullerene size[67]. Additionally, topological indices such as the first-moment pentagon signature P_1 and the second-moment hexagon signature H_2 have also been used to pinpoint the most stable isomers from a set of IPR-compliant isomers for a given fullerene size[23,68]. However, to the best of our knowledge, a generalized rule for estimating the stability, electronic properties, and solubility across the entire spectrum of potential fullerene isomers of varying sizes has not yet been established. Leveraging the comprehensive dataset computed in our study, we will examine the E_b, E_g, and logP values with various topological features, indices, and geometric measures.

To characterize the chemical environment of carbon atoms, C–C bonds, pentagons, and hexagons distributed across the fullerene cage, we introduced 4 types of topological features, as described in the "Methods" section. Linear regression has been used to fit E_b, E_g, and logP using these topological features (Fig. 5 and Table 2). We observed that the atom, bond, and hexagon features exhibit substantial linear correlations with E_b, resulting in R^2 values exceeding 0.96 (Fig. 5a, b, d). For example, the linear fitting between E_b and atom

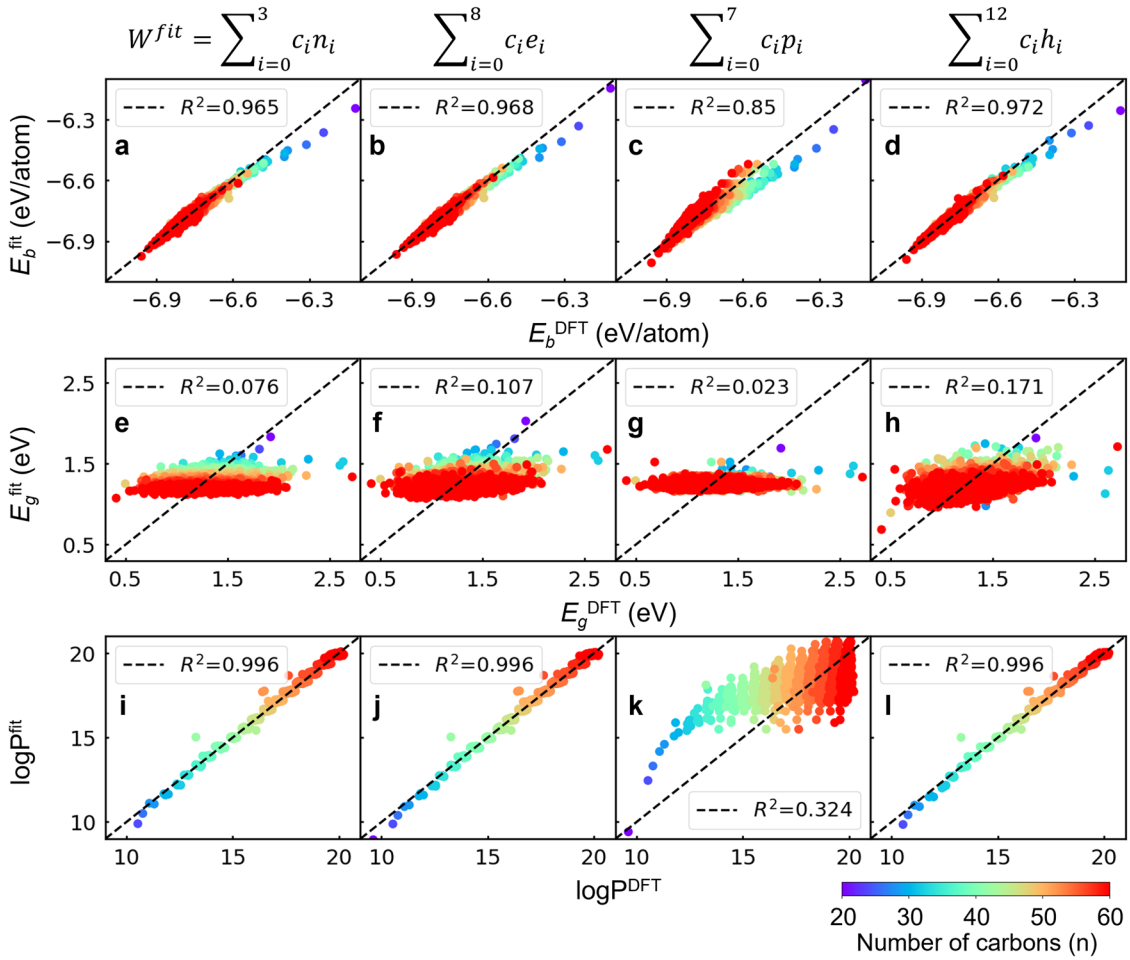

Fig. 5 | Linear fitting between topological features and fundamental properties of fullerenes. The parity plot of DFT-calculated binding energy (E_b (eV/atom)) (left panel), HOMO-LUMO gap (E_g (eV)) (middle panel), and ODCB–water partition coefficient (logP) (right panel) of all 5770 C_{20}–C_{60} fullerenes versus predicted values of linear regression based on **a**–**c** atom features, **d**–**f** bond features, **g**–**i** pentagon features, and **j**–**l** hexagon features. The color bar indicates varying fullerene sizes.

Table 2 | Coefficient of determination (R^2) obtained from linear fitting of topological features and geometric measures with fundamental properties, i.e., binding energy (E_b (eV/atom)), HOMO-LUMO gap (E_g (eV)), and ODCB–water partition coefficient (logP) for the 5770 C_{20}–C_{60} database

	R^2		
Topological features	E_b	E_g	logP
Atom features ($\{n_i \mid i = 0, 1, 2, 3\}$)	0.965	0.076	0.996
Bond features ($\{e_i \mid i = 0, \ldots, 8\}$)	0.968	0.107	0.996
Pentagon features ($\{p_i \mid i = 0, \ldots, 7\}$)	0.85	0.023	0.324
Hexagon features ($\{h_i \mid i = 0, \ldots, 12\}$)	0.972	0.171	0.996
Geometric measures			
Average bond length ($\bar{d}$)	0.946	0.036	0.545
Fowler asymmetry parameter (F_{asym})	0.092	0.006	0.042
Average curvature ($\bar{\kappa}$)	0.839	0.061	0.919
Volume (V)	0.721	0.062	0.977
Surface area (A)	0.595	0.067	0.995
Ratio A/V	0.878	0.054	0.85
Deviation from isoperimetric quotient (D_{IPQ})	0.707	0.006	0.194

features, i.e., n_0, n_1, n_2, and n_3, can be expressed as follows:

$$E_b\ (eV/atom) = c_0 n_0 + c_1 n_1 + c_2 n_2 + c_3 n_3 + b \tag{7}$$

We found that c_0 equals −4.9 meV/atom, indicating that the incorporation of an additional carbon atom fused by three hexagons (thereby increasing n_0 by 1) into fullerene cage will negligibly decrease E_b by 4.9 meV/atom. This result aligns with the fact that the growth of sp^2 carbons in a honeycomb lattice is thermodynamically favorable[69–71]. We have also noted that the coefficients c_1, c_2, and c_3 are positive and proportional to each other, with a correlation of $c_1 = \frac{1}{2}c_2 = \frac{1}{3}c_3$. Additionally, c_1 is approximately 13 orders of magnitude larger than the absolute value of c_0. Consequently, the linear fitting formula can be adjusted as follows:

$$E_b\ (eV/atom) = c_1\left(\frac{c_0}{c_1}n_0 + n_1 + 2n_2 + 3n_3\right) + b \tag{8}$$

Given that n_1, n_2, and n_3 denote the count of carbon atoms fused to 1, 2, and 3 adjacent pentagons, respectively, E_b will increase proportionally with the number of carbon atoms fused in pentagons, suggesting that the substitution of a hexagon with a pentagon will significantly diminish the thermodynamic stability of fullerene structures. Furthermore, the relationship of $c_1 = \frac{1}{2}c_2 = \frac{1}{3}c_3$ implies that as multiple pentagons fuse together (denoted by n_2 and n_3), E_b will increase dramatically, with the slope doubling or tripling. These findings are consistent with the IPR concept that fullerene

Table 3 | The binding energies (E_b) of 15 large-size fullerenes (Cn) calculated with B3LYP-D3/6-311G* compared to the values predicted by linear fitting models with atom, bond, pentagon, and hexagon features respectively across C_{20}–C_{60} fullerenes, together with various statistical measures, i.e., the mean deviation (MD), the overall mean absolute deviation (MAD), the standard deviation (SD), and the maximal absolute error (MAX)

Cn	E_b^{DFT}	E_b^{atom}	E_b^{bond}	$E_b^{pentagon}$	$E_b^{hexagon}$
C_{70}	−6.997	−7.027	−7.016	−6.988	−7.03
C_{72}	−6.98	−7.036	−7.023	−6.988	−7.028
C_{74}	−6.997	−7.046	−7.037	−6.988	−7.047
C_{76}	−7.005	−7.056	−7.045	−6.988	−7.056
C_{78}	−7.007	−7.066	−7.054	−6.988	−7.066
C_{80}	−7.01	−7.076	−7.063	−6.988	−7.077
C_{82}	−7.018	−7.086	−7.073	−6.988	−7.083
C_{84}	−7.008	−7.095	−7.081	−6.988	−7.071
C_{86}	−7.019	−7.105	−7.093	−6.988	−7.097
C_{90}	−7.036	−7.125	−7.109	−6.988	−7.099
C_{92}	−7.024	−7.135	−7.119	−6.988	−7.112
C_{94}	−7.027	−7.144	−7.129	−6.988	−7.119
C_{96}	−7.044	−7.154	−7.138	−6.988	−7.12
C_{98}	−7.028	−7.164	−7.148	−6.988	−7.123
C_{100}	−7.05	−7.174	−7.156	−6.988	−7.121
	MD	−0.082	−0.069	0.029	−0.067
	MAD	0.082	0.069	0.03	0.067
	SD	0.03	0.029	0.018	0.017
	MAX	0.135	0.12	0.062	0.095

The units for all the values are given in eV/atom. The molecular structures of these large-size fullerenes are shown in Fig. S7.

isomers with isolated pentagons tend to be more stable. Our interpretations of the values of c_0, c_1, c_2, and c_3 contribute to a deeper comprehension of the relationship between structure and stability in fullerene molecules ranging from C_{20} to C_{60}. In contrast, the pentagon features display a lower R^2 of 0.85 (Fig. 5c). The slightly inferior fit from pentagon features stems from their dependence on cage size. This is evidenced by the results obtained from analyzing only 1812 C_{60} isomers, where the R^2 value improved to 0.92 with pentagon features (Fig. S6 and Table S2).

To evaluate the transferability of the linear fitting model–defined as its capacity to learn and apply the correlations between topological features and binding energy in the dataset of C_{20}–C_{60} fullerenes to larger ones–we utilized the linear fitting equations derived from the C_{20}–C_{60} dataset to predict the E_b values of 15 larger IPR-conforming fullerenes (C_{70}–C_{100}, excluding C_{88}) sourced from Yoshida's Fullerene Library (see Fig. S7)[72]. The predictions obtained from atom, bond, and hexagon features demonstrate a consistent trend of decreasing E_b values with the increase in fullerene cage size, with E_b values closely matching the results from B3LYP-D3/6-311G* calculations (Table 3). This consistency underscores the robust transferability of the linear model fitted by the C_{20}–C_{60} dataset. When applying the linear model based on pentagon features, we observed that it assigned identical E_b values to these 15 IPR-conforming fullerenes. In essence, pentagon features fail to distinguish the stability of IPR-conforming fullerene molecules, as they all share the same p_0 value of 12 with other pentagon-related features (p_1 to p_7) set to 0 (Table S6). Such findings highlight the limitations of relying solely on pentagon features, as well as the IPR, for capturing the nuanced chemical environments of fullerenes across a range of sizes and topologies.

According to the previous study, the first-moment pentagon signature P_1, the first and second-moment hexagon signatures H_1 and H_2 were highly correlated with the stability of C_{60} fullerenes[23]. We further examined these three signatures for the extended dataset from C_{20} to C_{60}, and broadening the properties from binding energy to include E_g and logP (Table S7 and Fig. S8). Notably, within the group of fullerenes with identical size, a robust linear correlation emerges between topological signatures and the binding energies of fullerenes (Fig. S9), aligning with findings from previous reports[23]. For example, the linear relationship between E_b and P_1 in 1812 C_{60} database is described as:

$$E_b\ (eV/atom) = -6.9583 + 0.0167\,P_1 \tag{9}$$

The E_b value of C_{60}–#1 predicted by linear fitting is −6.9583 eV/atom as P_1 equals 0, closely matching the DFT-calculated value of −6.96 eV/atom. Introducing each pentagon-pentagon fusion (thereby incrementing P_1 by 1) results in an energy penalty of 0.0167 eV/atom (0.385 kcal/(mol · atom)) to E_b, equivalent to approximately 23.1 kcal/mol for C_{60} fullerenes. Our results are in good agreement with the value of 20–25 kcal/mol reported by Sure et al.[23]. However, it is essential to note that this linear correlation does not extend consistently across fullerenes with differing sizes (Fig. S8a–c).

Unlike E_b, E_g shows no linear relationship with both topological features (Fig. 5e–h and Table 2) and topological signatures (Fig. S8d–f and Table S7). The R^2 values for the correlation between the E_g and topological features fall below 0.2. This weak correlation still persists when only C_{60} isomers are considered (Table S2 and Fig. S6e–h), suggesting that linear models based on topological indices cannot effectively predict the electronic properties of fullerenes.

Implied by the correlation between E_b and logP, the solubility of fullerenes also exhibits a strong linear relationship with the atom, bond, and hexagon features, sharing an identical R^2 value of 0.996 (Table 2 and Fig. 5i, j, l). However, when exclusively considering C_{60} isomers, the respective R^2 values drop to 0.297, 0.353, and 0.448 (Table S2 and Fig. S6i, j, l). These findings underscore the notable relationship between logP values and fullerene cage size, which is well captured by topological features. Yet, within fullerene of the same size, logP values do not show a linear correlation with these topological features. Regarding pentagon features, the low R^2 value of 0.324 for logP in C_{20}–C_{60} dataset (Fig. 5k) indicates a limited ability of pentagon features to correlate with the fullerene cage size. This is consistent with the fact that the number of pentagons in a fullerene molecule remains fixed at 12, regardless of cage size, while the counts of carbon atoms, C–C bonds, and hexagons increase with cage size. This weak correlation with logP persists when focusing exclusively on C_{60} (Fig. S6k).

We further evaluated the relationships between logP and topological signatures. As illustrated in Fig. S8g–i, the hexagon-derived signatures, H_1 and H_2, display a stronger correlation with logP compared to the pentagon-based signature P_1 across the C_{20}–C_{60} dataset, demonstrating that hexagon features are more adept at capturing the properties of fullerenes across different sizes than pentagon features. Nevertheless, among fullerenes of the same size, logP values do not exhibit a linear correlation with any of the topological signatures (see Fig. S9g–i).

It is noted that the sum of pentagon features equals 12 for any given fullerene structure ($\sum_{i=0}^{7} p_i = 12$), while atom, bond, and hexagon features are associated with the carbon count. For example, the sum of atom features equals N for a fullerene structure with N carbon atoms ($\sum_{i=0}^{3} n_i = N$). When fitting fullerenes of varying sizes (carbon count N ranging from 20 to 60), $\{n_0, n_1, n_2, n_3\}$ are linearly independent. However, for a given fullerene size, such as the C_{60} dataset with 1812 structures, the atom features $\{n_0, n_1, n_2, n_3\}$ are linearly dependent due to the sum $\sum_{i=0}^{3} n_i = 60$ being a constant.

Correlations of geometric measures with fundamental properties

We also investigated a series of geometric measures, as described in "Methods," and their correlations with fundamental properties. The

coefficient of determination R^2 from linear fitting between geometric measures and fundamental properties can be found in Table 2. Across the entire fullerene dataset, $\bar{d}$, $\bar{\kappa}$, and A/V ratio exhibit strong linear correlations with fullerene binding energy E_b, with R^2 values of 0.946, 0.839, and 0.878, respectively (Fig. S10a, c, f). Other geometric measures, including F_{asym}, V, A, and D_{IPQ}, which assess the sphericity of the cage, exhibit a dependence on the fullerene size, as evident from the clustering of E_b values across various fullerene sizes (Fig. S10b, d, e, g). For example, when considering only the 1812 C_{60} isomers, the R^2 values of E_b fitted with F_{asym}, V, A, and D_{IPQ} are 0.714, 0.85, 0.627, and 0.875, respectively (Table S2 and Fig. S11b, d, e, g), consistent with previous findings[23]. These findings align with the observation that fullerene cages with larger volumes and more spherical shapes undergo reduced strain, leading to enhanced stability. However, it is crucial to note that this linear relationship does not extend uniformly across groups of fullerenes with differing sizes. When considering the entire fullerene dataset, these R^2 values decrease to 0.092, 0.721, 0.595, and 0.707 (Table 2 and Fig. S10b, d, e, g).

In contrast to E_b, E_g displays no correlation with all the geometric measures collected in this study, yielding R^2 values below 0.1 (Table 2 and Fig. S12). Importantly, this low correlation is independent of the cage size of fullerenes (see Table S2 and Fig. S13).

Additionally, we calculated the R^2 values for logP. As shown in Table 2 and Fig. S14, across the entire C_{20}–C_{60} fullerene database, both V and A exhibit strong correlations with logP values of fullerenes, yielding R^2 values of 0.977 and 0.995, respectively. Given that the V and A of a fullerene cage depend on its size, these findings align with our earlier discussion that logP strongly correlates with fullerene cage size. Interestingly, the R^2 values of both V versus logP and A versus logP for 1812 C_{60} isomers significantly decrease to 0.32 and 0.064, as presented in Table S2 and Fig. S15, suggesting that V and A are inadequate to differentiate the solubility of fullerenes with identical size. $\bar{\kappa}$ and A/V ratio also exhibit high correlates with logP values, resulting in R^2 values of 0.919 and 0.85 for C_{20}–C_{60}, respectively (see Fig. S14c, f). The observed similarity in the distributions of logP when plotted against $\bar{\kappa}$ and the A/V ratio suggests a strong correlation between A/V and $\bar{\kappa}$. In fact, $A/V = 3\bar{\kappa}$ for a perfect sphere.

In summary, the stability of fullerenes with different carbon atoms can be inferred from topological attributes such as atom, bond, and hexagon features, alongside geometric measures like $\bar{d}$, $\bar{\kappa}$, and A/V ratio. Furthermore, there is a notable correlation between the logP values across ODCB and water phases and the cage sizes of fullerenes. While these topological attributes and geometric measures (i.e., V and A) effectively illustrate trends from C_{20} to C_{60} fullerenes, they fall short in precisely predicting logP values for fullerenes of identical sizes. Nonetheless, accurately determining the E_g through these topological features and geometric measures remains a challenge. Future work might benefit from employing more complex models beyond linear regression or incorporating additional properties, to enhance the prediction of E_g values in fullerenes. This falls beyond the scope of our current paper.

To conclude, in this study, we built up a comprehensive computational dataset for C_{20}–C_{60} fullerenes with 5770 structures and computed 12 fundamental properties at DFT level. We conducted statistical analysis on stability, electronic properties, and solubility, and assessed the Pearson correlation coefficients among them. The observed weak correlations between E_g and both E_b ($r = 0.13$) and logP ($r = -0.24$) suggest that favorable electronic properties can be independently adjusted for particular applications without compromising stability and solubility. To gain a deeper understanding of the structure-property relationships in fullerenes, we introduced various multi-dimensional topological features, as well as geometric measures to further explore the nature and origin of these properties expanding beyond the commonly used IPR. Among them, atom, bond, and hexagon features prove effective in capturing the intricate local structural environments of the carbon atoms on the spherical cage. This capability facilitates precise predictions of fullerene stability across various sizes from C_{20} to C_{60} and ensures robust transferability to larger fullerene sizes beyond C_{60}. In contrast, pentagon features, along with the IPR, demonstrate inferior capacity due to the fixed number of 12 pentagons in a fullerene molecule, regardless of the cage size. The solubility of fullerenes represented by logP values adopts a high correlation with the fullerene cage size, which is well captured by atom, bond, and hexagon features. However, within fullerenes of the same size, logP values do not show a linear correlation with the topological features. Unlike E_b and logP, E_g exhibits no linear relationship with either topological features or geometric measures, leaving its correlation with fullerene structures as an open question. Our study lays a fundamental basis for future advancements in the functionalization and practical applications of fullerenes in the fields of energy conversion and nanomaterials sciences.

Methods

Dataset construction

We constructed the fullerene dataset from C_{20} to C_{60} with a total of 5770 structures. Table S8 lists the number of possible isomers, hexagon rings, C–C bonds, and IPR isomers for each fullerene cage size. The initial structures of fullerenes from C_{20} to C_{52} with 1249 structures were obtained from the online fullerene database[73], and 1812 isomer structures of C_{60} were adopted from the published paper[23]. For the remaining 2709 fullerene structures from C_{54} to C_{58}, we utilized the program FULLERENE (version 4.5) to generate the initial xyz coordinates and applied a force field specifically designed for fullerenes to optimize them[74]. The construction of the C_{54}–C_{58} dataset facilitates understanding the dependence of 12 fundamental properties on fullerene cage sizes ranging from C_{20} to C_{60}. All the optimized structures from our study can be found in the Supporting Information.

DFT calculations

All the DFT calculations were performed with Gaussian 16 package[75]. The B3LYP hybrid functional[76–78], together with D3 dispersion correction[79] and 6-31G* basis set, were employed for geometry optimization of all 5770 fullerene isomers. The maximum force tolerance was set to 0.02 eV/Å. To further confirm that the optimized structures are equilibrium geometries, the calculation of harmonic frequencies is essential. However, the complexity of these calculations[80,81] makes them impractical for the high-throughput screening studies we conducted. Nevertheless, given the initial structures are derived from the inherent symmetry of the fullerene cage, it suggests that the optimized structures are unlikely to resemble the transition state geometries. Subsequently, single-point calculations at singlet state were carried out with B3LYP functional and 6-311G* basis set to determine the energies and electronic properties. The continuum solvation model based on density (SMD) was used to obtain the solvation energy[82]. The dielectric constants of water and ODCB solvents are set as 78 and 10, respectively. In addition to the B3LYP functional, we have evaluated six other exchange-correlation (xc) functionals: CAM-B3LYP[83], LC-ωPBE[84], M06[85], M06-2X[85], PBE1PBE[86], and ωB97XD[87], to calculate E_b, E_g, and logP. As shown in Fig. S16, the B3LYP functional shows high correlation with all tested functionals in terms of E_b and logP, achieving R^2 values close to 1, whereas for E_g, B3LYP demonstrates lower correlation with range-separated xc functionals (such as CAM-B3LYP, LC-ωPBE, and ωB97XD) compared to other hybrid functionals (such as M06, M06-2X, and PBE1PBE). Previous study showed that, when benchmarked against the CCSD(T) method, the B3LYP functional demonstrates greater accuracy for fullerene isomerization energies compared to other xc functionals with higher exact Hartree-Fock exchange percentage[88]. Thus, we applied the B3LYP functional for all the DFT calculations in this study.

Topological features and indices

We introduced 4 types of topological features, namely, atom, bond, pentagon, and hexagon features, to characterize the structure of fullerenes. Each carbon atom in the fullerene cage can form three C–C bonds, and each bond is surrounded by four polygonal rings that belong to either pentagon or hexagon. The atom features categorize carbon atoms into 4 types, denoted as $\{n_i | i = 0, 1, 2, 3\}$, representing the count of carbon atoms fused to i adjacent

pentagons (Fig. S17). Similarly, the bond features define 9 types of C–C bonds $\{e_i|i = 0, \ldots, 8\}$ according to the number and arrangements of the 4 surrounding polygonal rings (Fig. S18). The pentagon features define 8 types of pentagons $\{p_i|i = 0, \ldots, 7\}$ depending on the number of arrangements of the 5 abutting polygons (Fig. S19). The hexagon features define 13 types of hexagons $\{h_i|i = 0, \ldots, 12\}$ which rely on the number and arrangements of the 6 neighboring polygons (Fig. S20). Table S9 lists the values of these 4 topological features for C_{20}–#1, C_{60}–#1, and C_{60}–#1812. It is important to note that Fowler and Manolopoulos also designed pentagon $\{p'_i|i = 0, \ldots, 5\}$ and hexagon $\{h'_i|i = 0, \ldots, 6\}$ indices; however, they only considered the number of pentagons (hexagons) attached to i other pentagons (hexagons) but ignored different arrangements of the pentagons (hexagons)[68,89]. For instance, p'_2 equals the sum of p_2 and p_3, whereas p'_3 equals the sum of p_4 and p_5 (see Fig. S19). They further consolidated the pentagon and hexagon indices into a more practical set of topological indices, namely, the pentagon signature P_1, and the n-th moment hexagon signature H_n using the formulas below:

$$P_1 = \frac{1}{2}\sum_{i=0}^{5} ip'_i \tag{10}$$

$$H_n = \sum_{i=0}^{6} i^n h'_i \tag{11}$$

For example, the IPR-conforming C_{60}-Ih buckminsterfullerene (C_{60}-#1), where all the pentagons are isolated by the hexagons, yields $p'_0 = 12$ and $h'_3 = 20$ with all other terms equating to zero, resulting in $P_1 = 0$, $H_1 = 60$, and $H_2 = 180$.

Geometric measures

We computed 7 geometric measures based on the optimized fullerene structures: average bond length ($\bar{d}$), volume (V), surface area (A), ratio of surface area and volume (A/V), Fowler asymmetry parameter (F_{asym})[90], deviation from isoperimetric quotient (D_{IPQ})[91], and the average curvature ($\bar{\kappa}$) of the fullerene cage.

$\bar{d}$ is the mean of all C–C bond lengths within the structure. V of a fullerene is determined by treating the cage as a polyhedron, composed entirely of pentagons and hexagons as its facets. Analogous to a polygon being segmented into triangles, the polyhedron fullerene cage can similarly be deconstructed into a collection of tetrahedra, with its volume being the sum of these tetrahedra. Similarly, A of a fullerene cage is the sum of the areas of the pentagons and hexagons constituting the fullerene cage, each of which can be further divided into triangles.

F_{asym} is calculated as follows:

$$F_{asym} = \sum_{i}^{N} \frac{(R_i - R_{av})^2}{R_{av}^2} \tag{12}$$

in which R_i is the radial distance of each carbon atom i from the center of mass of the fullerene, R_{av} is the average radial distance, and N represents the number of carbon atoms in the fullerene. F_{asym} equals 0 when all atoms lie on an ideal spherical fullerene cage, such as C_{60}–Ih (C_{60}–#1).

D_{IPQ} is calculated from the volume and surface area as follows:[23]

$$D_{IPD} = 1 - \frac{36\pi V^2}{A^3} \tag{13}$$

D_{IPD} is dimensionless and equals 0 for a perfect sphere.

The curvature at each carbon atom is defined as follows: in a given fullerene molecule, a carbon atom, designated as the central carbon, forms a tetrahedron with its three neighboring atoms. This tetrahedron uniquely dictates a circumsphere where all four carbon atoms reside. The radius of the circumsphere represents the curvature radius (R) at the central carbon atom's position. Consequently, the curvature (κ) is the reciprocal of R:

$$\kappa = \frac{1}{R} \tag{14}$$

Thus, $\bar{\kappa}$ of the fullerene cage can be obtained by calculating the radius of curvature at each carbon atom of the fullerene cage.

Data availability

All the optimized C_{20}–C_{60} structures (XYZ files) and the calculated fundamental properties in a.csv file, both are provided in the supporting information.

Received: 11 March 2024; Accepted: 1 September 2024;
Published online: 28 September 2024

References

1. Ramar, V. & Balraj, A. Critical review on carbon-based nanomaterial for carbon capture: technical challenges, opportunities, and future perspectives. *Energy Fuels* **36**, 13479–13505 (2022).
2. Speranza, G. Carbon nanomaterials: synthesis, functionalization and sensing applications. *Nanomaterials* **11**, 967 (2021).
3. Zhang, B.-T., Zheng, X., Li, H.-F. & Lin, J.-M. Application of carbon-based nanomaterials in sample preparation: a review. *Anal. Chim. Acta* **784**, 1–17 (2013).
4. Scida, K., Stege, P. W., Haby, G., Messina, G. A. & García, C. D. Recent applications of carbon-based nanomaterials in analytical chemistry: critical review. *Anal. Chim. Acta* **691**, 6–17 (2011).
5. Maiti, D., Tong, X., Mou, X. & Yang, K. Carbon-based nanomaterials for biomedical applications: a recent study. *Front. Pharmacol.* **9**, 1401 (2019).
6. Thurston, W. P. Shapes of polyhedra and triangulations of the sphere. Preprint at https://arxiv.org/abs/math/9801088 (1998).
7. Paukov, M., Kramberger, C., Begichev, I., Kharlamova, M. & Burdanova, M. Functionalized fullerenes and their applications in electrochemistry, solar cells, and nanoelectronics. *Materials* **16**, 1276 (2023).
8. Chen, J.-X. et al. Using fullerene fragments as acceptors to construct thermally activated delayed fluorescence emitters for high-efficiency organic light-emitting diodes. *J. Chem. Eng.* **435**, 134731 (2022).
9. Kausar, A., Ahmad, I., Maaza, M., Eisa, M. & Bocchetta, P. Polymer/fullerene nanocomposite for optoelectronics-moving toward green technology. *J. Compos. Sci.* **6**, 393 (2022).
10. Jakšić, J. et al. Exploring fullerene derivatives for optoelectronic applications: synthesis and characterization study. *Phys. Chem. Chem. Phys.* **26**, 517–523 (2024).
11. Yin, H., Lin, H., Zong, Y. & Wang, X.-D. The recent advances in C_{60} micro/nanostructures and their optoelectronic applications. *Org. Electron.* **93**, 106142 (2021).
12. Ganesamoorthy, R., Sathiyan, G. & Sakthivel, P. Fullerene based acceptors for efficient bulk heterojunction organic solar cell applications. *Sol. Energy Mater. Sol. Cells* **161**, 102–148 (2017).
13. Dennler, G., Scharber, M. C. & Brabec, C. J. Polymer-fullerene bulk-heterojunction solar cells. *Adv. Mater.* **21**, 1323–1338 (2009).
14. Brabec, C. J. et al. Polymer–fullerene bulk-heterojunction solar cells. *Adv. Mater.* **22**, 3839–3856 (2010).
15. Agrawal, P. S., Belkhode, P. N., Brijpuriya, D. S., Gouda, S. P. & Rokhum, S. L. Stimulation in fullerene for adsorbing pollutant gases: a review. *Chem. Phys. Impact* **6**, 100156 (2023).
16. Haghgoo, S. & Nekoei, A.-R. Metal oxide adsorption on fullerene C_{60} and its potential for adsorption of pollutant gases; density functional theory studies. *RSC Adv.* **11**, 17377–17390 (2021).
17. Castro, E., Garcia, A. H., Zavala, G. & Echegoyen, L. Fullerenes in biology and medicine. *J. Mater. Chem. B* **5**, 6523–6535 (2017).

18. Yáñez-Sedeño, P., Campuzano, S. & Pingarrón, J. M. Fullerenes in electrochemical catalytic and affinity biosensing: a review. *C* **3**, 21 (2017).
19. Pilehvar, S. & De Wael, K. Recent advances in electrochemical biosensors based on fullerene-C_{60} nano-structured platforms. *Nanocarbons for Electroanalysis*, 173–196 (Wiley, 2017).
20. Colt, J. R. & Scuseria, G. E. An ab initio study of the C_{78} fullerene isomers. *Chem. Phys. Lett.* **199**, 505–512 (1992).
21. Wang, Y., Díaz-Tendero, S., Alcamí, M. & Martín, F. Cage connectivity and frontier π orbitals govern the relative stability of charged fullerene isomers. *Nat. Chem.* **7**, 927–934 (2015).
22. Aghajamali, A. & Karton, A. Can force fields developed for carbon nanomaterials describe the isomerization energies of fullerenes? *Chem. Phys. Lett.* **779**, 138853 (2021).
23. Sure, R., Hansen, A., Schwerdtfeger, P. & Grimme, S. Comprehensive theoretical study of all 1812 C_{60} isomers. *Phys. Chem. Chem. Phys.* **19**, 14296–14305 (2017).
24. Sabirov, D. S., Ori, O. & László, I. Isomers of the C_{84} fullerene: a theoretical consideration within energetic, structural, and topological approaches. *Fuller. Nanotub. Carbon Nanostructures* **26**, 100–110 (2018).
25. Austin, S., Fowler, P., Orlandi, G., Manolopoulos, D. & Zerbetto, F. Relative stabilities of C_{76} isomers. A numerical test of the fullerene isolated-pentagon rule. *Chem. Phys. Lett.* **226**, 219–225 (1994).
26. Song, X., Mao, R., Wang, Z. & Qi, J. Structural and spectral properties of a non-classical C_{58} isomer and its fluorinated derivatives in theory. *RSC Adv.* **11**, 1472–1481 (2021).
27. Mao, R., Wang, Z., Song, X., Chen, W.-K. & Qi, J. Structural and spectral properties of a nonclassical C_{66} isomer with its hydrogenated derivative $C_{66}H_4$ in theory. *ACS Omega* **6**, 27101–27111 (2021).
28. Cui, Y.-H., Chen, D.-L., Tian, W. Q. & Feng, J.-K. Structures, stabilities, and electronic and optical properties of C_{62} fullerene isomers. *J. Phys. Chem. A* **111**, 7933–7939 (2007).
29. Zhao, D., Liu, S., Rong, C., Zhong, A. & Liu, S. Toward understanding the isomeric stability of fullerenes with density functional theory and the information-theoretic approach. *ACS Omega* **3**, 17986–17990 (2018).
30. Chan, B. & Karton, A. Computational insights into the singlet–triplet energy gaps, ionization energies, and electron affinities for a diverse set of 812 small fullerenes (C_{20}–C_{50}). *Phys. Chem. Chem. Phys.* **25**, 10899–10906 (2023).
31. Shin, H. et al. Cohesion energetics of carbon allotropes: quantum Monte Carlo study. *J. Chem. Phys*. **140**, 114702 (2014).
32. Ke, S.-H., Baranger, H. U. & Yang, W. Addition energies of fullerenes and carbon nanotubes as quantum dots: the role of symmetry. *Phys. Rev. Lett.* **91**, 116803–116803 (2003).
33. Kabir, M., Mukherjee, S. & Saha-Dasgupta, T. Substantial reduction of stone-wales activation barrier in fullerene. *Phys. Rev. B* **84**, 205404 (2011).
34. Dresselhaus, M. S., Dresselhaus, G. & Eklund, P. C. Science of fullerenes and carbon nanotubes: their properties and applications (Elsevier, 1996).
35. Liakos, D. G., Sparta, M., Kesharwani, M. K., Martin, J. M. & Neese, F. Exploring the accuracy limits of local pair natural orbital coupled-cluster theory. *J. Chem. Theory Comput.* **11**, 1525–1539 (2015).
36. Minenkov, Y., Chermak, E. & Cavallo, L. Accuracy of DLPNO–CCSD (T) method for noncovalent bond dissociation enthalpies from coinage metal cation complexes. *J. Chem. Theory Comput.* **11**, 4664–4676 (2015).
37. Prinzbach, H. et al. Gas-phase production and photoelectron spectroscopy of the smallest fullerene, C_{20}. *Nature* **407**, 60–63 (2000).
38. Piskoti, C., Yarger, J. & Zettl, A. C_{36}, a new carbon solid. *Nature* **393**, 771–774 (1998).
39. Xie, S.-Y. et al. Capturing the labile fullerene [50] as $C_{50}Cl_{10}$. *Science* **304**, 699–699 (2004).
40. Tan, Y.-Z. et al. Two I_h-symmetry-breaking C_{60} isomers stabilized by chlorination. *Nat. Mater.* **7**, 790–794 (2008).
41. Mumyatov, A. V. & Troshin, P. A. A review on fullerene derivatives with reduced electron affinity as acceptor materials for organic solar cells. *Energies* **16**, 1924 (2023).
42. Thompson, B. C. & Fréchet, J. M. Polymer–fullerene composite solar cells. *Angew. Chem. Int. Ed. Engl.* **47**, 58–77 (2008).
43. Kooistra, F. B. et al. New C_{84} derivative and its application in a bulk heterojunction solar cell. *Chem. Mater.* **18**, 3068–3073 (2006).
44. Scharber, M. C. et al. Design rules for donors in bulk-heterojunction solar cells—towards 10% energy-conversion efficiency. *Adv. Mater.* **18**, 789–794 (2006).
45. Eibeck, A. et al. Predicting power conversion efficiency of organic photovoltaics: models and data analysis. *ACS Omega* **6**, 23764–23775 (2021).
46. Hachmann, J. et al. The Harvard clean energy project: large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.* **2**, 2241–2251 (2011).
47. Gueymard, C. A., Myers, D. & Emery, K. Proposed reference irradiance spectra for solar energy systems testing. *Sol. Energy* **73**, 443–467 (2002).
48. Scharber, M. C. & Sariciftci, N. S. Efficiency of bulk-heterojunction organic solar cells. *Prog. Polym. Sci.* **38**, 1929–1940 (2013).
49. Grancharov, G. et al. Flexible polymer–organic solar cells based on P3HT: PCBM bulk heterojunction active layer constructed under environmental conditions. *Molecules* **26**, 6890 (2021).
50. Krebs, F. C. Fabrication and processing of polymer solar cells: a review of printing and coating techniques. *Sol. Energy Mater. Sol. Cells* **93**, 394–412 (2009).
51. Arndt, A. P., Gerhard, M., Koch, M., Lemmer, U. & Howard, I. A. Identifying charge-transfer states in polymer: fullerene heterojunctions by their emission polarization anisotropy. *J. Phys. Chem. C* **121**, 6357–6364 (2017).
52. Arkhipov, V., Heremans, P. & Bässler, H. Why is exciton dissociation so efficient at the interface between a conjugated polymer and an electron acceptor? *Appl. Phys. Lett.* **82**, 4605–4607 (2003).
53. de Gier, H. D., Jahani, F., Broer, R., Hummelen, J. C. & Havenith, R. W. Promising strategy to improve charge separation in organic photovoltaics: installing permanent dipoles in PCBM analogues. *J. Phys. Chem. A* **120**, 4664–4671 (2016).
54. Carsten, B. et al. Examining the effect of the dipole moment on charge separation in donor–acceptor polymers for organic photovoltaic applications. *J. Am. Chem. Soc.* **133**, 20468–20475 (2011).
55. Yang, Y. et al. Reversible fullerene electrochemistry: correlation with the HOMO-LUMO energy difference for C_{60}, C_{70}, C_{76}, C_{78}, and C_{84}. *J. Am. Chem. Soc.* **117**, 7801–7804 (1995).
56. Semenov, K. N. et al. Solubility of light fullerenes in organic solvents. *J. Chem. Eng. Data* **55**, 13–36 (2010).
57. Kharissova, O. V., Oliva González, C. M. & Kharisov, B. I. Solubilization and dispersion of carbon allotropes in water and non-aqueous solvents. *Ind. Eng. Chem. Res.* **57**, 12624–12645 (2018).
58. Ferreira, R. A., dos Santos, J. R., Falleiros, J. M., da Silva, E. L. & dos Santos, J. D. Functionalization of fullerene isomers [60, 70, 80 and 82] with R-thiazolidinethione: a DFT study. *Chem. Phys. Lett.* **769**, 138421 (2021).
59. Hinkle, K. R. & Phelan Jr, F. R. Solvation of carbon nanoparticles in water/alcohol mixtures: using molecular simulation to probe energetics, structure, and dynamics. *J. Phys. Chem. C* **121**, 22926–22938 (2017).
60. Varanasi, S., Guskova, O., John, A. & Sommer, J.-U. Water around fullerene shape amphiphiles: a molecular dynamics simulation study of hydrophobic hydration. *J. Chem. Phys*. **142**, 224308 (2015).
61. Athawale, M. V., Jamadagni, S. N. & Garde, S. How hydrophobic hydration responds to solute size and attractions: theory and simulations. *J. Chem. Phys*. **131**, 115102 (2009).

62. Muthukrishnan, A. & Sangaranarayanan, M. Hydration energies of C_{60} and C_{70} fullerenes—a novel Monte Carlo simulation study. *Chem. Phys.* **331**, 200–206 (2007).
63. Graziano, G. On the pairwise hydrophobic interaction of fullerene. *Chem. Phys. Lett.* **499**, 79–82 (2010).
64. Stukalin, E. B., Korobov, M. V. & Avramenko, N. V. Solvation free energies of the fullerenes C_{60} and C_{70} in the framework of polarizable continuum model. *J. Phys. Chem. B* **107**, 9692–9700 (2003).
65. Bannan, C. C., Calabró, G., Kyu, D. Y. & Mobley, D. L. Calculating partition coefficients of small molecules in octanol/water and cyclohexane/water. *J. Chem. Theory Comput.* **12**, 4015–4024 (2016).
66. Ben-Naim, A. & Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **81**, 2016–2027 (1984).
67. Kroto, H. The stability of the fullerenes C_n, with $n = 24, 28, 32, 36, 50, 60$ and 70. *Nature* **329**, 529–531 (1987).
68. Fowler, P. W., Caporossi, G. & Hansen, P. Distance matrices, Wiener indices, and related invariants of fullerenes. *J. Phys. Chem. A* **105**, 6232–6242 (2001).
69. Karfunkel, H. R. & Dressler, T. New hypothetical carbon allotropes of remarkable stability estimated by MNDO solid-state SCF computations. *J. Am. Chem. Soc.* **114**, 2285–2288 (1992).
70. Balaban, A., Klein, D. & Folden, C. Diamond-graphite hybrids. *Chem. Phys. Lett.* **217**, 266–270 (1994).
71. Park, N. & Ihm, J. Electronic structure and mechanical stability of the graphitic honeycomb lattice. *Phys. Rev. B* **62**, 7614 (2000).
72. https://nanotube.msu.edu/fullerene/fullerene-isomers.html.
73. https://houseofgraphs.org/meta-directory/fullerenes.
74. Schwerdtfeger, P., Wirz, L. & Avery, J. Program fullerene: a software package for constructing and analyzing structures of regular fullerenes. *J. Comput. Chem.* **34**, 1508–1526 (2013).
75. Frisch, M. et al. Gaussian16. Revision C.01. Gaussian Inc. (Wallingford, CT 2016).
76. Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372–1377 (1993).
77. Stephens, P. J., Devlin, F. J., Chabalowski, C. F. & Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **98**, 11623–11627 (1994).
78. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785 (1988).
79. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys*. **132**, 154104 (2010).
80. Jing, D. & Pan, Z. Molecular vibrational modes of C_{60} and C_{70} via finite element method. *Eur. J. Mech. A/Solids* **28**, 948–954 (2009).
81. Schettino, V., Pagliai, M., Ciabini, L. & Cardini, G. The vibrational spectrum of fullerene C_{60}. *J. Phys. Chem. A* **105**, 11192–11196 (2001).
82. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).
83. Yanai, T., Tew, D. P. & Handy, N. C. A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP). *Chem. Phys. Lett.* **393**, 51–57 (2004).
84. Henderson, T. M., Izmaylov, A. F., Scalmani, G. & Scuseria, G. E. Can short-range hybrids describe long-range-dependent properties? *J. Chem. Phys*. **131**, 044108 (2009).
85. Zhao, Y. & Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **120**, 215–241 (2008).
86. Adamo, C. & Barone, V. Toward reliable density functional methods without adjustable parameters: the PBE0 model. *J. Chem. Phys.* **110**, 6158–6170 (1999).
87. Chai, J.-D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008).
88. Karton, A. Fullerenes pose a strain on hybrid density functional theory. *J. Phys. Chem. A* **126**, 4709–4720 (2022).
89. Fowler, P. W. & Manolopoulos, D. E. *An Atlas of Fullerenes* (Courier Corporation, 2007).
90. Fowler, P. W., Heine, T. & Zerbetto, F. Competition between even and odd fullerenes: C_{118}, C_{119}, and C_{120}. *J. Phys. Chem. A* **104**, 9625–9629 (2000).
91. Pisanski, T., Kaufman, M., Bokal, D., Kirby, E. C. & Graovac, A. Isoperimetric quotient for fullerenes and other polyhedral cages. *J. Chem. Inf. Comput.* **37**, 1028–1032 (1997).

Acknowledgements

We acknowledge Zhiyu Wang for the help on CCSD(T) calculations. This work was financially supported by the University of Florida's new faculty start-up funding. The authors acknowledge the University of Florida Research Computing for providing computational resources and support that have contributed to the research results reported in this publication.

Author contributions

B.L. and M.L. initiated this study. B.L. was responsible for conducting the theoretical calculations and analyzing the data. The Python code used to compute the topological features and geometric measures was developed by J.J. The manuscript was written by B.L. and M.L., with all authors contributing through discussions.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-024-01410-7.

Correspondence and requests for materials should be addressed to Mingjie Liu.