Structural prediction of carbon cluster isomers with machine-learning potential

H. D. Nguyen[†]

Faculty of Physics, University of Science, Vietnam National University, Hanoi, 334 Nguyen Trai, Thanh Xuan, Hanoi 100000, Vietnam

E-mail: [†]huydn@hus.edu.vn

Abstract. *Structural prediction of low-energy isomers of carbon twelve-atom clusters is carried out using the recently developed machine-learning potential GAP-20. The GAP-20 agrees with density-functional theory calculations regarding geometric structures and average C-C bond lengths for most isomers. However, the GAP-20 substantially lowers the energies of cage-like structures, resulting in a wrong ground state. A comparison of the cohesive energies with the density-functional theory points out that the GAP-20 only gives good results for monocyclic rings. Two multicyclic rings appear as new low-energy isomers, which have yet to be discovered in previous research.*

1. Introduction

A system of carbon atoms is ubiquitous in daily life. Carbon clusters, for example, are widely distributed in the atmosphere, crust, and biology, forming a variety of morphologies. For decades, the synthesis of novel carbon allotropes has attracted inclusive interest. In the 80s, small carbon molecules in the shape of chains are first observed [1]. By atomic manipulation of oxide molecules, generation of cyclocarbon is reported [2, 3]. Previous pioneering experiments lead to the discovery of C_{60} fullerene, a new form of carbon clusters with icosahedral symmetry [4]. To open an avenue for elusive carbon-rich materials with exotic properties, a comprehensive understanding of carbon clusters' geometric/electronic structure is essential.

Based on the vaporization of carbon-rich targets, some experimental techniques such as arc discharge, laser ablation, or chemical vapor deposition [5–7] have allowed the in-lab synthesis of carbon clusters, C_n. In addition, mass spectrometry confirms their formation for various n up to 500, or even at $n \approx 8000$ [8–10]. Due to the competition between surface and bulk, carbon clusters exhibit various structural motifs and geometrical frustration. The energy landscape is complex, and the number of possible isomers increases rapidly with the number of atoms. Even with modern probe techniques, obtaining unambiguous structural information from experiments is often tricky. Therefore, the theoretical investigation, which can be divided into two main categories, first-principles techniques and molecular dynamics simulation, is needed.

Density-functional theory (DFT) calculations performed by Shi and co-workers predict 20 isomers of C_{11} (C_{12}) with a global minimum corresponding to a monocyclic ring [11]. Relying on the coupled cluster method, Manna and Martin have shed light on the stability of some C_{20} and C_{24} isomers, finding that C_{24} is the smallest fullerene [12]. Employing the diffusion Monte Carlo approach, Cleland and colleagues examine electron correlation effects, which reordered relative energies in ring, bowl, and cage isomers of the C_{20} molecule [13]. However, the demanding computational costs of such highly predictive frameworks restrict the cluster datasets up to a few tens of atoms.

On the other hand, molecular dynamics substantially reduces the calculation overheads. Using Brenner potential, Kosimov and coworkers find that the lowest-energy configurations of C_n consist of cyclocarbons for $11 \leq n \leq 18$ and graphitic nanoflakes for $19 \leq n \leq 55$ [14]. Based on structures obtained with Brenner potential, Mauney *et al.* subsequently carry out DFT calculations and report that the ground-state configurations of C_n are monocyclic rings for $10 \leq n \leq 23$ and mostly fullerenes for $n \geq 24$ [15]. Since the empirical potentials are often developed for a given carbon material, their transferability to systems of different morphologies is not guaranteed. For C_n of up to 200 atoms, Karasulu and colleagues show that empirical potentials generally predict different stable structures and cohesive energies compared to the DFT data [16].

Machine-learning frameworks have opened an alternative way in the progress of potential development. Based on an extensive and accurate reference database generated with DFT calculations, the procedure typically involves training the energies and forces and a subsequent interpolation to predict new atomic environments. After successful training, quantum-mechanical accuracy can be achieved with machine learning potential (MLP) at a reduced computational cost. Developed by Rowe *et al.*, the recent MLP for carbon, named GAP-20, has shown excellent accuracy in predicting formation energies and phonon dispersions of numerous carbon allotropes [17]. Qian *et al.* also find that the GAP-20 accurately models the thermal stabilities, defect energies, and van der Waals interactions for crystalline and amorphous carbon [18]. Despite being successfully applied to various carbon phases, studies on the transferability of the GAP-20 to cluster morphologies are very limited [16].

This study employs the GAP-20 to find low-energy isomers of C_{12}. The current results reveal that the GAP-20 gives structures similar to the DFT in most cases. The average C-C bond lengths obtained by the two methods agree well, and the errors are less than 3.5%. The isomer relative/cohesive energy is inaccurately predicted by the GAP-20, particularly for cage-like structures. Ring-like isomers tend to be energetically preferable, and cage-like isomers prefer triangles and squares. Two low-energy rings not discovered in the previous rings are found and discussed.

2. Computational Procedures

The C_{12} allotropes are initially obtained by exploring the energy landscape with particle swarm optimization. Implemented in CALYPSO software [19, 20], this method has proven superior to previous techniques for structural prediction, such as simulated annealing or genetic algorithms [21]. The swarm attains collective intelligence by continuously updating a population of new isomers at each iteration/generation, leading to an efficient search. The procedure will continue until the convergence criterion is reached. In this study, the search is carried out five times, and the population and iteration/generation numbers in each run are 30 and 200, respectively. For a generated coordinate, geometry optimization is performed with the GAP-20 via the QUIP module in LAMMPS [22]. Combining the conjugate gradient algorithm and FIRE damped-dynamics [23] followed by Hessian-free truncated Newton algorithm, the optimization protocol ensures the energy and force tolerance criteria of 10^{-12} eV and 10^{-8} eV/Å, respectively, and the pressure contribution to the enthalpy is negligible ($\leq 10^{-5}$ eV).
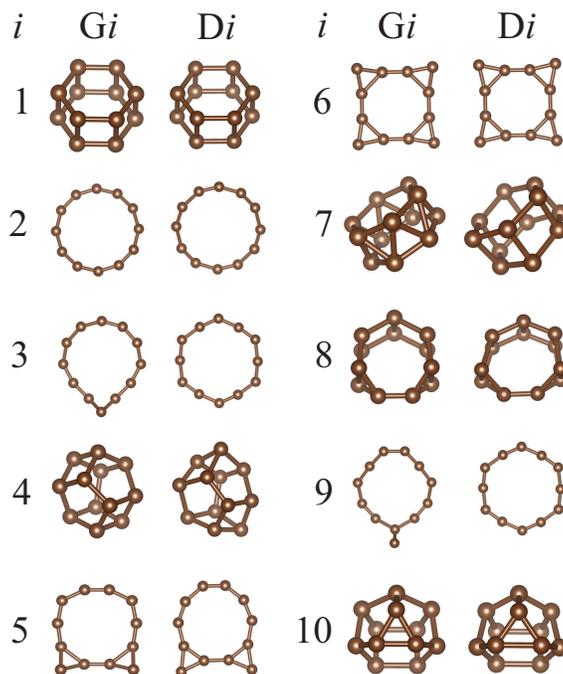
Fig. 1. (Color online) Optimized structures of the lowest-energy C_{12} isomers obtained by the GAP-20, Gi (left), and the corresponding structures after further optimization with the DFT, Di (right), with i from 1 to 10.

The obtained structures using LAMMPS are then further optimized by DFT calculations, as integrated in Quantum ESPRESSO package [24, 25]. The electron-ion interaction is described by the standard solid state pseudopotential for carbon [26] using the projector-augmented wave method [27]. Since the dispersion correction gives an insignificant contribution to geometry optimization [28], the electron-electron interaction is treated with the PBE version of the generalized gradient approximation [29]. Plane-wave basis sets are chosen with a kinetic energy cutoff for

wavefunctions (charge density) of 612 (4896) eV. The imposed periodic boundary conditions are in three directions, and the unit cell includes at least 10 Å of vacuum to minimize the spurious interactions between periodic images of the cluster. Geometry optimization is carried out using BFGS quasi-newton algorithm until the force components on each atom are below 0.05 eV/Å.

3. Results and Discussion

3.1. Low-energy isomers

To evaluate the structural search performance using current MLP, we show in Figure 1 the optimized structures of 10 lowest-energy isomers found by the GAP-20, Gi, and their corresponding structures after further optimization by the DFT, Di, with i from 1 to 10. The index i increases as the total energy of the Gi isomer increases. Various structural types are observed, including cages ($i = 1, 4, 7, 10$), monocyclic rings ($i = 2, 3, 9$), multicyclic rings ($i = 5, 6$), and others ($i = 8$). Cage-like isomers mostly contain hexagons, pentagons, or squares, and some triangles for the G7 and G10 isomers. Monocyclic rings include cumulenic (G2), deformed (G3), and tadpole (G9) rings, and multicyclic systems consist of two or more three-membered rings.

The structures obtained by the two calculation methods are somewhat similar. Noticeable structural changes occur for the G3, G7, and G9 configurations. DFT relaxation transforms the deformed and tadpole rings into hexagonal and polyynic rings (D3 and D9 are the same structure). From the G7 isomer, two hexagons buckle and rotate about each other, removing all triangles. Overall, the GAP-20 provides a suitable starting geometry for subsequent DFT optimization.

Table 1. Point group symmetry PGS, average C-C bond length d_{C-C} (Å), and singlet-triplet splitting E_{S-T} (eV) of the low-energy C_{12} isomer found by the GAP-20 Gi, and of the corresponding DFT optimized structure Di, with i from 1 to 10.

i	PGS		d_{C-C}			E_{S-T}
	Gi	Di	Gi	Di	Error	Di
1	D_{6h}	D_{6h}	1.425	1.475	-3.4%	0.172
2	D_{12h}	D_{6h}	1.293	1.300	-0.5%	-0.085
3	C_s	C_{6h}	1.310	1.302	0.6%	0.159
4	D_{2d}	C_2	1.435	1.483	-3.2%	0.544
5	C_s	C_{2v}	1.362	1.361	0.1%	0.601
6	C_{4v}	D_{4h}	1.397	1.410	-0.9%	1.196
7	D_{6d}	D_{3d}	1.455	1.479	-1.6%	0.496
8	C_{2v}	C_s	1.409	1.437	-2.0%	-0.445
9	C_s	C_{6h}	1.321	1.302	1.5%	0.159
10	C_s	C_s	1.422	1.470	-3.3%	0.937

The point-group symmetries and the average C-C bond lengths of the C_{12} isomers are summarized in Table 1. The average bond lengths obtained by the GAP-20 are very close to those calculated by the DFT, with errors of less than 3.5%. The absolute errors are the smallest for ring-like isomers and are the largest for cage-like isomers. The GAP-20 fails to capture the Jahn-Teller distortion, which causes the G2 cumulene, i.e., same bond lengths, to become the D2 polyyne,

i.e., alternating bond lengths. The polyynic D2 (D3) without (with) bond-angle alternation exhibits D_{6h} (C_{6h}) symmetry and alternating single bonds of 1.37 Å (1.35 Å) and triple bonds of 1.24 Å (1.26 Å). The existence of polyynic structures without and with bond-angle alternation has also been observed in experiments for the cyclo[18]carbon [2, 3]. The GAP-20 gives reasonable average bond length for the monocyclic rings, multicyclic rings, and cages, corresponding to the sp-, sp^2-, and sp^3-hybridized molecular carbon allotrope, respectively.
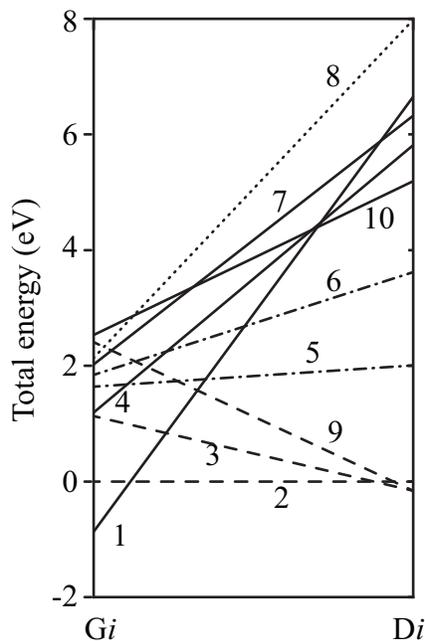
Fig. 2. Total energies of the low-energy C_{12} isomers obtained by the GAP-20 and DFT. Values on the left (right) represents the energies relative to the energy of the G2 (D2) isomer calculated by the GAP-20 (DFT). Values are connected by lines representing the structural type: solid, cages; dashed, monocyclic rings; dash-dotted, multicyclic rings; dotted, others. The number next to a line indicates the isomer index i.

Figure 2 presents the total energies of the C_{12} isomers relative to the energy of the G2 (D2) configuration calculated by the GAP-20 (DFT). Current MLP generally cannot produce a correct ordering and an acceptable approximation (≤ 0.05 eV/atom for fullerenes as in Ref. [17]) to isomer energies compared to the DFT results. For small clusters, the formation of cages should be unfavorable due to strain and curvature. The GAP-20, however, strongly underestimates the energies of the cage-like isomers (solid lines). The underestimation is the most severe for the G1 configuration, rendering it the ground state. In addition, the MLP faces limitations in describing the energetics of the deformed and tadpole rings. DFT calculations reveal that the G3 and G9 structures are unstable and will transform to the D3 ground-state configuration (dashed lines). The finding of the hexagonal and polyynic ring as the lowest-energy structure agrees well with those in previous studies [11, 30]. We note that the GAP-20 can not obtain the D3 as a stable isomer, as geometry optimization starting from the D3 will eventually converge to the G2. The energy

landscapes obtained by the two calculation methods are essentially different for the deformed and tadpole rings. The performance of the MLP can be improved by adding these configurations to the database for the re-training procedure. On the contrary, the energy difference is small for multicyclic rings with several three-membered components (dash-dotted lines), suggesting that employment of the GAP-20 may be suitable for such cases.

From the DFT energy picture, forming ring-like structures is energetically favorable for small clusters. Although previous literature agrees that the monocyclic ring is the ground-state configuration [15, 16, 31], it is unclear if the D2 polyyene is found. In addition, the energy difference between the D_{6h} and C_{6h} polyynic structures (D2 and D3) is tiny (~ 0.16 eV), implying that they are challenging to differentiate as in experiments of the cyclo[18]carbon. Besides the ground-state structure, the D1, D4, and D7 isomers correspond to the C40, C36, and C39 structures observed by Shi *et al.* [11]. The D5 and D6 configurations with the relative energies of 2.0 and 3.6 eV above the ground-state energy are also two new isomers compared to the previous work [11]. The small relative energies imply that these isomers may exist in the experiment procedures. The C_{12} cages will likely welcome triangles and squares over pentagons or hexagons, frequently observed in fullerenes. Interestingly, the D10 configuration with one triangle has the lowest energy among the cages.

To discuss the chemical stability, singlet-triplet splitting (E_{S-T}), presented in Tab. 1, is an important feature. Defined as the energy of the triplet state minus that of the singlet state, the negative (positive) sign in E_{S-T} implies that the triplet (singlet) state is more stable. In general, the larger the magnitude of the E_{S-T} the smaller the reactivity of the cluster. The splitting can be determined experimentally using anion photoelectron spectroscopy [32, 33]. It is noted that DFT does not adequately treat the singlet state for open-shell systems. However, Shirazi *et al.* have pointed out that the average error in E_{S-T} is just 3.54 kcal/mol (0.14 eV) compared to the highly accurate coupled cluster calculation for some carbenes [28]. The use of DFT may still provide reasonable results. As seen in Tab. 1, monocyclic rings have the smallest E_{S-T} of -0.09 and 0.16 eV, indicating that these isomers are highly reactive and may be hard to be isolated. On the contrary, the multicyclic ring D6 and cage-like structure D10 have the most significant E_{S-T} of -1.20 and -0.94 eV, respectively. Given the high energy of the cage-like D10, the D6 may appear as a kinetically stable isomer under experimental conditions.

3.2. *Cohesive energies*

Figure 3 presents the cohesive energies of the C_{12} clusters. The cohesive energy E_c is computed as

$$E_c = \frac{E_{tot} - nE_{ref}}{n}, \tag{1}$$

where E_{tot} is the total energy of a cluster, predicted by the DFT or GAP-20, n is equal to 12, and E_{ref} is the reference energy of an isolated carbon atom. E_{ref} equals 0 by construction for the machine-learned potential. Except for the G7, the differences in DFT energies of the unoptimized (Gi-DFT) and optimized (Di-DFT) structures are small, implying that local minima in the energy landscape of the GAP-20 are close to those of the DFT. However, current MLP only gives a good agreement of E_c for monocyclic rings, even though the G3 and D3 (G9 and D9) structures are quite different. The differences in E_c are significant for cages and G8 isomer, and the GAP-20 does not give satisfactory results.
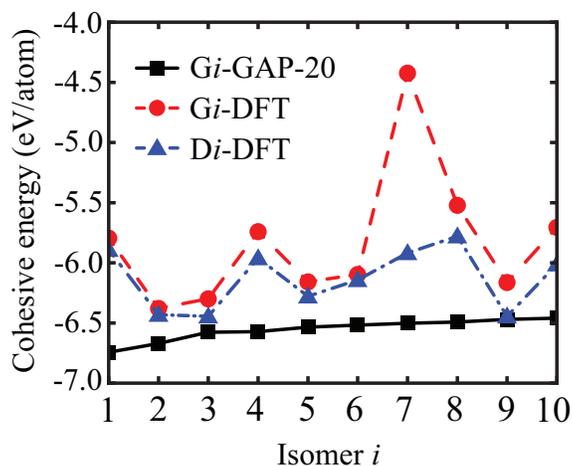
Fig. 3. (Color online) Cohesive energies (eV/atom) of the G*i* isomer calculated by the GAP-20 (G*i*-GAP-20 – black solid line) and the DFT (G*i*-DFT – red dashed line), and those of the D*i* isomer calculated by the DFT (D*i*-DFT – blue dash-dotted lines).

4. Conclusions

This study evaluates the performance of the GAP-20 for the structural search of C_{12} cluster isomers. The structures obtained by the GAP-20 and the DFT are relatively similar in most cases. The computed average C-C bond lengths using the two methods are close to each other, with less than 3.5% errors. However, the GAP-20 incorrectly describes the relative/cohesive energies of the isomers. Cage-like structures are strongly stabilized, leading to a wrong ground-state configuration. Starting from the GAP-20 optimized coordinates, subsequent DFT relaxations find two multicyclic rings as new low-energy isomers not found in the previous study. In future research, retraining of the GAP-20 will be carried out, and the stabilities of the C_{12} clusters will be investigated by performing a thermodynamic study or vibrational analysis.

References

[1] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl, and R. E. Smalley, *C_{60}: Buckminsterfullerene*, Nature **318** (1985) 162.

[2] K. Kaiser, L. M. Scriven, F. Schulz, P. Gawel, L. Gross, and H. L. Anderson, *An sp-hybridized molecular carbon allotrope, cyclo[18]carbon*, Science **365** (2019) 1299.

[3] L. M. Scriven, K. Kaiser, F. Schulz, A. J. Sterling, S. L. Woltering, P. Gawel *et al.*, *Synthesis of Cyclo[18]carbon via Debromination of $C_{18}Br_6$*, J. Am. Chem. Soc. **142** (2020) 12921.

[4] H. W. Kroto, *The stability of the fullerenes C_n, with n = 24, 28, 32, 36, 50, 60 and 70*, Nature **329** (1987) 529.

[5] E. Barborini, P. Piseri, A. Li Bassi, A. C. Ferrari, C. E. Bottani, and P. Milani, *Synthesis of carbon films with controlled nanostructure by separation of neutral clusters in supersonic beams*, Chem. Phys. Lett. **300** (1999) 633.

[6] C. Lifshitz, *Carbon clusters*, Int. J. Mass Spectrom. **200** (2000) 423.

[7] A. Hu, Q. B. Lu, W. W. Duley and M. Rybachuk, *Spectroscopic characterization of carbon chains in nanostructured tetrahedral carbon films synthesized by femtosecond pulsed laser deposition*, J. Chem. Phys. **126** (2007) 154705.

[8] H. Kietzmann, R. Rochow, G. Ganteför, and W. Eberhardt, *Electronic Structure of Small Fullerenes: Evidence for the High Stability of C_{32}*, Phys. Rev. Lett. **81** (1998) 5378..

[9] H. Shinohara, H. Sato, Y. Saito, A. Izuoka, T. Sugawara, H. Ito *et al.*, *Extraction and mass spectroscopic charac-terization of giant fullerences up to C_{500}*, Rapid Commun. Mass Spectrom. **6** (1992) 413.

[10] J. Houška, N. R. Panyala, E. M. Peña-Méndez, and J. Havel, *Mass spectrometry of nanodiamonds*, Rapid Com-mun. Mass Spectrom. **23** (2009) 1125.

[11] L. T. Shi, Z. Q. Wang, C. E Hu, Y. Cheng, J. Zhu, and G. F. Ji, *Possible lower energy isomer of carbon clusters C_n (n = 11, 12) via particle swarm optimization algorithm: Ab initio investigation*, Chem. Phys. Lett. **721** (2019) 74..

[12] D. Manna and J. M. L. Martin, *What Are the Ground State Structures of C_{20} and C_{24}? An Explicitly Correlated Ab Initio Approach*, J. Phys. Chem. A **120** (2016) 153.

[13] D. M Cleland, E. K Fletcher, A. Kuperman, and M. C Per, *Electron correlation effects in isomers of C_{20}*, J. Phys. Mater. **3** (2020) 025006.

[14] D. P. Kosimov, A. A. Dzhurakhalov, and F. M. Peeters, *Carbon clusters: From ring structures to nanographene*, Phys. Rev. B **81** (2010) 195414.

[15] C. Mauney, M. B. Nardelli, and D. Lazzati, *Formation and Properties of Astrophysical Carbonaceous Dust. I. Ab-Initio Calculations of the Configuration and Binding Energies of Small Carbon Clusters*, ApJ **800** (2015) 30.

[16] B. Karasulu, J. M. Leyssale, P. Rowe, C. Weber, and C. de Tomas, *Accelerating the prediction of large carbon clusters via structure search: Evaluation of machine-learning and classical potentials*, Carbon **191** (2022) 255.

[17] P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi, and A. Michaelide, *An accurate and transferable machine learning potential for carbon*, J. Chem. Phys. **153** (2020) 034702.

[18] C. Qian, B. McLean, D. Hedman, and F. Ding, *A comprehensive assessment of empirical potentials for carbon materials*, APL Mater. **9** (2021) 061102.

[19] Y. Wang, J. Lv, L. Zhu, and Y. Ma, *Crystal structure prediction via particle-swarm optimization*, Phys. Rev. B **82** (2010) 094116.

[20] J. Lv, Y. Wang, L. Zhu, and Y. M. Ma, *Particle-swarm structure prediction on clusters*, J. Chem. Phys. **137** (2012) 084104. https://doi.org/10.1063/1.4746757

[21] G. Jana, A. Mitra, S. Pan, S. Sural, and P. K. Chattaraj, *Modified Particle Swarm Optimization Algorithms for the Generation of Stable Structures of Carbon Clusters, C_n (n = 3–6, 10)*, Front. Chem **7** (2019) 485.

[22] S. Plimpton, *Fast Parallel Algorithms for Short-Range Molecular Dynamics*, J. Chem. Phys. **117** (1995) 1.

[23] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch, *Structural Relaxation Made Simple*, Phys. Rev. Lett. **97** (2006) 170201.

[24] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni *et al.*, *QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials*, J. Phys. : Condens. Matter **21** (2009) 395502.

[25] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra *et al.*, *Advanced capabilities for materials modelling with Quantum ESPRESSO*, J. Phys. : Condens. Matter **29** (2017) 465901.

[26] G. Prandini, A. Marrazzo, I. E. Castelli, N. Mounet, and N. Marzari, *Precision and efficiency in solid-state pseu-dopotential calculations*, Npj Comput. Mater. **4** (2018) 72.. http://materialscloud.org/sssp

[27] P. E. Blöchl, *Projector augmented-wave method*, Phys. Rev. B **50** (1994) 17953.

[28] R. G. Shirazi, D. A. Pantazis, and F. Neese, *Performance of density functional theory and orbital-optimised second-order perturbation theory methods for geometries and singlet–triplet state splittings of aryl-carbenes*, Mol. Phys. **118** (2020) e1764644.

[29] J. P. Perdew, K. Burke, and M. Ernzerhof, *Generalized Gradient Approximation Made Simple*, Phys. Rev. Lett. **77** (1996) 3865.

[30] J. M. L. Martin, J. El-Yazal, and J.-P. Frarqois, *Structure and vibrational spectra of carbon clusters C_n (n = 2–10, 12, 14, 16, 18) using density functional theory including exact exchange contributions*, Chem. Phys. Lett. **242** (1995) 570. https://doi.org/10.1016/0009-2614(95)00801-A

[31] T. W. Yen and S. K. Lai, *Use of density functional theory method to calculate structures of neutral carbon clusters C_n ($3 \leq n \leq 24$) and study their variability of structural forms*, J. Chem. Phys. **142** (2015) 084313. https://doi.org/10.1063/1.4908561

[32] C. Xu, G. R. Burton, T, R. Taylor, and D. M. Neumark, *Photoelectron spectroscopy of C_4^-, C_6^-, and C_8^-*, J. Chem. Phys. **107** (1997) 3428.

[33] A. V. Orden and R. J. Saykally, *Small Carbon Clusters: Spectroscopy, Structure, and Energetics*, Chem. Rev. **98** (1998) 2313.