Data in Brief 55 (2024) 110652

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

# Density functional tight-binding derived data of gas capture in functionalized carbophenes

Chad E. Junkermeier[a,b,*], Jedediah Kobebel[b], Kat Lavarez[a], R. Martin Adra[a], Jirui Yang[a], Valeria Aparicio Diaz[a], Ricardo Paupitz[c], George Psofogiannaki[d]

[a] Department of Physics and Astronomy, University of Hawai'i at Mānoa, Honolulu HI 96822, USA
[b] Science, Technology, Engineering, and Mathematics Department, University of Hawai'i Maui College, Kahului HI 96732, USA
[c] Departamento de Física, Universidade Estadual Paulista, UNESP, 13506-900, Rio Claro, SP, Brazil
[d] Department of Chemical and Biological Engineering, University of Ottawa, Ottawa, Ontario, Canada

A R T I C L E I N F O

*Article history:*
Received 22 November 2023
Revised 28 May 2024
Accepted 11 June 2024
Available online 17 June 2024

Dataset link: DFTB based models of gas molecules adsorbed in functionalized carbophenes (Original data)

*Keywords:*
Covalent organic framework
Porous materials
2-dimensional materials
Greenhouse gas adsorption
Hydrogen adsorption

A B S T R A C T

The presented data relates to the investigation of the adsorption properties of carbon dioxide ($CO_2$), methane ($CH_4$), and dihydrogen ($H_2$) within pristine and functionalized carbophene pores. The carbophenes were functionalized with one of the groups carboxyl (COOH), amine ($NH_2$), nitro ($NO_2$), hydroxyl (OH), or an amide ($CONH_2$, NHCOOH, and $N(COOH)_2$) groups. The systems were optimized using the density functional tight-binding theory code DFTB+ (precompiled Version 19.1) with the matsci Slater-Koster files on the Mana high performance computing cluster at the University of Hawai'i at Mānoa. The dataset consists of the molecular geometries, lattice vectors, and the total energies for each specific system. One possible use of the data is for training or validating force fields for running molecular dynamics simulations.



* Corresponding author.
*E-mail address:* junkerme@hawaii.edu (C.E. Junkermeier).

https://doi.org/10.1016/j.dib.2024.110652

Specifications Table

| | |
|---|---|
| Subject | Computational Materials Science |
| Specific subject area | This work focuses on materials that are two-dimensional and functionalizable. |
| Data format | Raw, analyzed |
| Type of data | Table |
| Data collection | Crystal structure calculations were performed using DFTB+ (pre-compiled Version 19.1), utilizing the *matsci* Slater-Koster files, and were collected into an edn database formatted file; executed on Mana, the University of Hawai‘i at Mānoa’s high performance computing cluster. Each record contained the DFTB+ calculated total energy, lattice vectors, atomic coordinates, and atomic charge. The records also included parameter values for the type of carbophene, the type of functional group, the position of the functional group (if any), the type of gas molecule (if any), and a record number. |
| Data source location | The data was generated using the Mana high performance computing cluster at the University of Hawai‘i at Mānoa. |
| Data accessibility | Repository name: Mendeley Data<br>Data identification number: 10.17632/bxkbbs2553.3<br>Direct URL to data: https://data.mendeley.com/datasets/bxkbbs2553/3 [1] |
| Related research article | C. E. Junkermeier, E. Larmand, J.-C. Morais, J. Kobebel, K. Lavarez, R. M. Adra, J. Yang, V. A. Diaz, R. Paupitz, G. Psofogiannakis, Functionalized carbophenes as high-capacity versatile gas adsorbents: An ab initio study, Computational Materials Science 232 (2024) 112665. [2] |

## 1. Value of the Data

- **Why are these data valuable?**
  - Carbophenes are a novel material that may have been synthesized a few years ago. The structural properties of carbophenes are not well known. The models provided in this data set will allow other researchers to check, refine, or build upon our theoretical work.
  - The dataset aims to provide insight into the adsorption properties of functionalized carbophenes depending on certain modifications and parameters, such as the desorption temperature being raised.
  - The adsorption properties can be compared to rival metal-based materials on atmospheric gas capturing and hydrogen storing to see if it provides better capability in these fields
  - These data display potential routes in modifying the adsorption energies of the gases, such as determining the best functional groups
- **Who can benefit from these data?**
  - Engineers who work for scientific organizations, such as NASA or NOAA, for projects like spaceships or submarines. Any project that would benefit from air purification in a confined space to pursue further research in a different scientific subject area or the same area as carbophene investigation.
  - Scientific communities interested in researching carbon-based **materials for** developing cleaner and more efficient energy sources.
- **How can these data be reused by other researchers?**
  - To further investigate this relatively new carbon-based material, which has not yet been physically developed, by evaluating functional groups for selective permeability or reference for multi-layered systems of this material.

## 2. Data Description

There are four files included in the data set. One file, example_DFTB_job_script.sh, is a shell script given to model how the data was created. The other three files containing the txt suffix contain the molecular data presented in the edn data scheme. In the data records, all energies are in eV, lattice vectors and atomic coordinates are in Å, and the charge is in electrons. edn is a

**Table 1**
Description of edn style key-value pairs contained in the molecular records found in *Gas_molecules_in_functionalized_carbophenes.txt. The* files *Gas_molecules.txt and Functionalized_carbophenes.txt* contain subsets of these key-value pairs.

| Key | Value type | Value Description | Example |
|---|---|---|---|
| :ad | string | States the name of the adsorbent gas molecule ($H_2$, $CO_2$, or $CH_4$). | :ad "CO2" |
| :adEn | real | The adsorption energy of a gas molecule in functionalized carbophene. | :adEn 0.10289999999898214 |
| :case | integer | A unique identifier for the results of each optimization. | :case 127 |
| :Ef | Real | Formation energy | :Ef 75.92850000000038 |
| :energy | real | The total energy of the system as computed by DFTB+. | :energy −13,327.3056 |
| :group | string | States the type of functional group used (pristine, COOH, $NH_2$, $NO_2$, OH, $CONH_2$, NHCOOH, or $N(COOH)_2$). | :group "COOH" |
| :lvs | Nested vectors | Lattice vectors of the 2D crystal structure. | :lvs [[46.46283972 26.82533302 0.0]<br>[46.46283972 −26.82533302 0.0]<br>[0.0 0.0 30]] |
| :mol | string | The molecular geometry where the first column is the atomic species, the following three columns are the atomic coordinates in Angstroms, and the last column is the number of valence electrons in the atom. | :mol "C 17.60541160 0.73150238 0.99999895 4.01509586<br>C 19.11032018 0.69179882 0.99999958 3.98714136<br>C 17.60245366 −0.77621085 0.99999912 4.01532815<br>. . . . .<br>. . . . .<br>. . . . .<br>H 58.65099926 −2.08283039 1.00003721 0.93733121<br>C 47.36188486 −10.50913873 0.99314751 3.39064889<br>O 47.36059376 −9.96943657 −0.35009573 6.39382556<br>O 47.35731272 −9.75789958 1.94926005 6.44792099<br>H 47.35923494 −9.01014265 −0.19709528 0.73789002<br>O 42.94406466 −8.26737736 0.89626370 6.35777856<br>C 43.99696224 −7.85536240 0.58199928 3.28473294<br>O 45.04939380 −7.44985084 0.25887939 6.35740253″ |
| :n | integer | Bonding site of the functional group | :n 21 |
| :rings | integer | States the type of N-carbophene or the linear N-phenylenes (this work focuses on 3-, 4-, and 5-carbophene) | :rings 5 |

general-purpose human readable extensible information data scheme designed to transfer data from one program to another with minimal textual noise [3]. edn records are a subset of the Clojure (and Clojurescript) programming language(s); thus, the key-value pairs of a record may contain any type of data. Table 1 presents a listing of the key-value pairs we are using.

The file *Gas_molecules.txt* contains the records of the relaxed molecular structures for the three gas molecules studied (i.e., $CO_2$, $CH_4$, and $H_2$) and two extra gas molecules used in computing the formation energy (i.e., $O_2$ and $N_2$). Each record contains key-value pairs for the keys :ad, :case, :energy, and :mol. Fig. 1 displays the record for a methane molecule in free space.

The file *Functionalized_carbophenes.txt* includes the data for each of the functionalized carbophenes crystal structures in the absence of any gas molecules. Each record contains key-value pairs for the keys :case, :energy, :lvs, :mol, :n, and :rings. Fig. 2 displays a partial record for a 5-carbophene with a nitro functional group. This file contains a total of 32 records.

The file *Gas_molecules_in_functionalized_carbophenes.txt* contains 721 records of a functionalized carbophene with an adsorbed gas molecule. Each record contains key-value pairs for the

```
{:ad "CH4"   :energy -86.3274 :case 2
    :mol  "C       5.00005179       5.00089453       5.00027110       4.20550473
    H       5.63690322       4.35877051       5.63744539       0.94863682
    H       5.64063634       5.63910608       4.36267317       0.94862250
    H       4.35839462       4.36434987       4.36211881       0.94856773
    H       4.36447404       5.64203900       5.63981153       0.94866821"}
```


**Fig. 1.** An example edn record from Gas_molecule.txt.


```
{:energy -13114.3808,
 :group "NO2",
 :n 19,
 :rings 5,
 :lvs
 [[46.46283972 26.82533302 0.0]
  [46.46283972 -26.82533302 0.0]
  [0.0 0.0 30]],
 :mol
 "C     17.60478788      0.73123539      1.00000000      4.01473619
   C     19.11099878      0.69218087      1.00000000      3.98744307
   C     17.60180525     -0.77475130      1.00000000      4.01510457
   .     .                .               .               .
   .     .                .               .               .
   .     .                .               .               .
   H     63.12825828     -4.69208079      1.00000000      0.93742719
   H     58.65008586     -2.08247259      1.00000000      0.93717190
   N     43.51662824    -10.58851248      0.99999996      4.21696196
   O     43.51267700     -9.98883793     -0.09298120      6.44992597
   O     43.51260411     -9.98883812      2.09298094      6.44992591",
 :case 23}
```


**Fig. 2.** An example edn record from Functionalized_carbophenes.txt.


```
{:adEn 0.02980000000037819,
 :rings 3,
 :group "NO2",
 :Ef 30.073699999999647,
 :n 7,
 :energy -6090.1404,
 :case 429,
 :mol
 "C      8.79624766      0.38426440      1.15535893      4.00960711
   C     10.30386272      0.35174837      1.15535896      3.98695122
   C      8.79482435     -1.11741435      1.15535900      4.01273909
   C     10.30334016     -1.08728804      1.15535899      3.98628262
   .      .                .               .               .
   .      .                .               .               .
   .      .                .               .               .
   H     26.70620300      1.71278626      1.15584794      0.93255409
   N     23.12531024     -4.23443074      1.15517930      4.21649973
   O     23.12553936     -3.63657236      0.06123646      6.45035756
   O     23.12612387     -3.63549462      2.24851154      6.45020502
   C     25.24158889     -0.88547548      0.29714089      4.20603735
   H     26.25016842     -0.47970961      0.50480987      0.95852018
   H     24.75370341     -0.27091179     -0.48304869      0.94793720
   H     25.33126705     -1.93031693     -0.05671501      0.94288589
   H     24.63308184     -0.86012437      1.22053186      0.94267144",
 :ad "CH4",
 :lvs
 [[23.32833952 13.46862311 0.0]
  [23.32818803 -13.46853564 0.0]
  [0.0 0.0 30.0]]}
```


**Fig. 3.** An example edn record from Gas_molecules_in_functionalized_carbophenes.txt.

keys :ad, :adEn, :case, :energy, :lvs, :mol, :n, and :rings. Fig. 3 displays a partial record for a 5-carbophene with a nitro functional group and an adsorbed methane molecule.

Fig. 4 presents pristine models of 3-, 4-, and 5-carbophene. In this dataset, we functionalized only one adsorption site at a time. By functionalization, we mean replaying an H atom with another functional group. In 3-carbophene and 4-carbophene, only one bonding site each is not equivalent under point group transformations. Meanwhile, in 5-carbophene, two bonding sites are not equivalent under point group transformations. In the records, the bonding site (:n) value is based on the position of the C atom bonding site in the list of atomic coordinates. Thus, for 3-carbophene, the replaced H atom is bonded to the seventh C atom in the list of coordinates. For 5-carbophene, we replace either the H atom connected to the 19th or 21st carbon in the coordinate list.

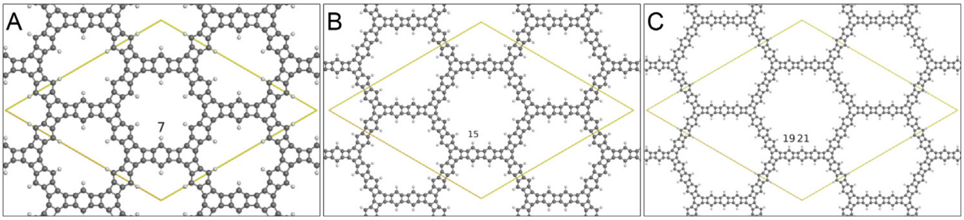


**Fig. 4.** 2-by-2 supercells of (A) 3-carbophene, (B) 4-carbophene, and (C) 5-carbophene. The gold lines outline the supercells, and the numbers 7, 15, 19, and 21 are located above the hydrogen atom that was replaced with a functional group when the respective value of the key :n showed up in a record.

## 3. Experimental Design, Materials and Methods

All data records were created using density functional-based tight binding (DFTB+) calculations to optimize the crystal structures and position of the adsorbed gas molecule [4–6]. The relaxed free space gas molecules (in *Gas_molecules.txt*) and optimized carbophene crystal structures (*Functionalized_carbophenes.txt*) were first computed separately. Model carbophene crystal structures were created by Junkermeier based on the discussion of Du et al. [7,8] Model functionalized carbophenes were created by Junkermeier et al. for a discussion of how the functional groups change the electronic structure of the carbophene crystals [9]. In the preceding works, Junkermeier et al. optimized the models using the *matsci Slater-Koster* files with Universal force field (UFF) dispersion corrections in DFTB+ [10,11]. In this work, all of the models from the preceding works were reoptimized in DFTB+ with the Grimme-D3 dispersion corrections in place of the UFF dispersion terms [12]. Hydrogen bond corrections were computed using the D3H5 method [13]. The functionalized carbophenes were combined with the $CO_2$, $CH_4$, and $H_2$ gas molecules into a set of input structures and relaxed using a conjugate gradient algorithm with a maximum force difference of $10^{-5}$ Ha/Bohr and an SCC maximum tolerance of $10^{-4}$ electrons as convergence criteria. The results are given in Gas_molecules_in_functionalized_carbophenes.txt.

To create the initial gas molecule-functionalized carbophene models, we first used the Greenwood library to place the gas molecule near the pore structure programmatically [14]. Using Greenwood, we placed molecules at the center of the pore or in one of several positions around where a functional group would be. Besides putting the gas molecules in the places just described, the gas molecules were also given a few different orientations at each position. Despite having different functional groups, the gas molecules were placed in the same positions and orientations relative to the functionalized carbophene crystal structure. The Greenwood script placed the crystal structure within a job (shell) script like example_DFTB_job_script.sh. This model-building method resulted in 621 job scripts being produced, but only 401 of those jobs ran correctly—most of the jobs that did not run correctly failed during the initiation of DFTB+. After reviewing the results of this method, we found that it mostly gave adsorption energies $<$ 0.1 eV. However, some jobs resulted in adsorption energies between 0.1 and 0.4 eV. The low adsorption energies were likely a result of the chosen relaxation method failing to move the gas molecules into better positions for physisorption. Because a large percentage of calculations ran incorrectly, and the low adsorption energy values obtained from those that ran, we determined that instead of investigating the jobs that ran incorrectly we would focus on methods that were likely to place the gas molecules in ideal positions. Thus, we manually created models using VMD [15,16]. Using VMD, we placed the gas molecules into positions and orientations we believed would lead to strong physisorption. Utilizing this method, we created 360 job scripts, 22 of which did not produce optimized geometries.

Upon the successful completion of a relaxation calculation, the following results were recorded: total energy (:energy), lattice vectors (:lvs), and molecular structure (:mol). Using the total energies recorded for the gas molecules, functionalized carbophene crystal structures, and the carbophene crystal structures with adsorbed gas molecules, we were able to compute the adsorption energies $E_{ads}$ of the gas molecules in the functionalized carbophenes. The adsorption energies listed in the edn records are calculated using

$$E_{ads} = E_{carbophene} + E_{gas} - E_{carbophene+gas},$$

where $E_{carbophene+gas}$ is the total energy of a relaxed functionalized carbophene and adsorbed gas molecule system, $E_{carbophene}$ is the total energy of the functionalized carbophene alone in the periodic cell, and $E_{gas}$ is the total energy of the gas molecule alone in the periodic cell. The resultant $E_{ads}$ value was then recorded as the value for the :adEn key of each record in *Gas_molecules_in_functionalized_carbophenes.txt*.

Each record in Gas_molecules_in_functionalized_carbophenes.txt also contains a key-value pair for the formation energy, :Ef. The formation energies are computed by

$$\boldsymbol{E_F} = \boldsymbol{E}_{carbophene+gas} - \boldsymbol{N}_C\left(\frac{\boldsymbol{E}_{gr}}{\boldsymbol{N}_{gr}}\right) - \boldsymbol{N}_H\left(\frac{\boldsymbol{E}_{H_2}}{\boldsymbol{2}}\right) - \boldsymbol{N}_N\left(\frac{\boldsymbol{E}_{N_2}}{\boldsymbol{2}}\right) - \boldsymbol{N}_O\left(\frac{\boldsymbol{E}_{O_2}}{\boldsymbol{2}}\right),$$

where $N_C$, $N_H$, $N_N$, and $N_O$ are respectively the number of carbon atoms, hydrogen atoms, nitrogen atoms, and oxygen atoms in the relaxed functionalized carbophene and adsorbed gas molecule system, $E_{gr}$, $E_{H_2}$, $E_{H_2}$, and $E_{O_2}$ are the total energies of possible precursors, namely a graphene unit cell and the gas molecules $N_2$, $H_2$, and $O_2$, and $N_{gr}$ is the number of atoms in the graphene unit cell.

## Limitations

We ran more gas adsorption optimization calculations than are recorded in Gas_molecules_in_functionalized_carbophenes.txt. Two types of errors occurred that resulted in the results not being recorded:

A calculation would end without generating the final geometry files; this happened 22 times out of 434 manually created input crystal models and 220 times out of 621 jobs created using Greenwood. As stated above, most calculations that did not produce results ended in the initialization stage of running DFTB+. A smaller proportion of the jobs failed to produce results because a calculation would end, and the total energy would be reported, but the resultant geometry files were not generated. These failures to produce results could be due to a problem in our models or to an instability of a compute node. While many problems with a compute node may cause jobs to fail, a recent update to the instructions for running on our cluster indicates that interactive jobs on a node may interfere with the running of submitted jobs and give new commands to be placed in job scripts to overcome this problem.

Eighteen of the manually generated geometries started with a $CO_2$ molecule that was triangular instead of linear. Based on the triangular $CO_2$, the relaxed geometries had 13 systems with physisorbed triangular $CO_2$ molecules and five with $CO_2$ chemisorbed. Because the input $CO_2$ geometries were not physical, we removed these 18 results.

## Ethics Statement

This work does not involve human or animal subjects and has not collected data from social media platforms. The authors confirm that they have read and followed the ethical requirements for publication.

## Data Availability

DFTB based models of gas molecules adsorbed in functionalized carbophenes (Original data) (Mendeley Data).

## CRediT Author Statement

**Chad E. Junkermeier:** Conceptualization, Data curation, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing; **Jedediah Kobebel:** Investigation, Writing – review & editing; **Kat Lavarez:** Investigation, Writing – review & editing; **R. Martin Adra:** Investigation, Writing – review & editing; **Jirui Yang:** Investigation, Writing – review & editing; **Ricardo Paupitz:** Writing – review & editing; **George Psofogiannaki:** Conceptualization, Writing – review & editing.

## Acknowledgements

C.E.J., K.L., R.M.A., and V.A.D. were supported by the United States funding agency NSF under award #2113011, and advanced computing resources from the University of Hawaii Information Technology Services – Cyberinfrastructure, funded in part by the National Science Foundation awards #2201428 and #2232862. J.Y. was supported by the United States funding agency NSF under award #2121848. Hawai'i Space Grant Consortium NASA Space Grant Undergraduate Traineeship supported J.K. R. Paupitz acknowledges Brazilian funding agencies: FAPESP (grant #2018/03961-5) and CNPq (grants #437034/2018-6, #310369/2017-7) and computing resources supplied by the Center for Scientific Computing (NCC/GridUNESP) of the São Paulo State University (UNESP).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

[1] C.E. Junkermeier, et al. DFTB based models of gas molecules adsorbed in functionalized carbophenes (2024), doi: 10.17632/BXKBBS2553.3.
[2] C.E. Junkermeier, et al., Functionalized carbophenes as high-capacity versatile gas adsorbents: an ab initio study, Comput. Mater. Sci. 232 (2024) 112665.
[3] R. Hickey, edn-format/edn: Extensible Data Notation, 2014. https://github.com/edn-format/edn.
[4] B. Aradi, B. Hourahine, Frauenheim, Th. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method, J. Phys. Chem A 111 (2007) 5678–5684.
[5] T. Kubar, et al., Parametrization of the SCC-DFTB Method for Halogens, J. Chem. Theory Comput. 9 (2013) 2939–2949.
[6] J. Frenzel, A.F. Oliveira, N. Jardillier, T. Heine, G. Seifert, Semi-relativistic, self-consistent charge Slater-Koster tables for density-functional based tight-binding (DFTB) for materials science simulations, Zeolites 2 (2004) 7.
[7] Q.-S. Du, et al., A new type of two-dimensional carbon crystal prepared from 1,3,5-trihydroxybenzene, Sci. Rep. 7 (2017) 40796.
[8] C.E. Junkermeier, J.P. Luben, R. Paupitz, N-Carbophenes: two-dimensional covalent organic frameworks derived from linear N-phenylenes, Mater. Res. Express 6 (2019) 115103.
[9] C.E. Junkermeier, G. Psofogiannakis, R. Paupitz, Covalent adsorption of functional groups on N-carbophenes, Mater. Res. Express (2022), doi:10.1088/2053-1591/ac4c19.
[10] B. Lukose, A. Kuc, J. Frenzel, T. Heine, On the reticular construction concept of covalent organic frameworks, Beilst. J. Nanotechnol. 1 (2010) 60.
[11] A.K. Rappe, C.J. Casewit, K.S. Colwell, W.A. Goddard, W.M. Skiff, UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations, J. Am. Chem. Soc. 114 (1992) 10024–10035.
[12] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. 132 (2010) 154104.
[13] J. Řezáč, Empirical Self-Consistent Correction for the Description of Hydrogen Bonds in DFTB3, J. Chem. Theory Comput. 13 (2017) 4804–4817.
[14] C.E. Junkermeier, Greenwood: a library for creating molecular models and processing molecular dynamics simulations (2018), https://github.com/cjunkermeier/greenwood.
[15] W. Humphrey, A. Dalke, K. Schulten, VMD – visual molecular dynamics, J. Mol. Graph 14 (1996) 33–38.
[16] J. Stone, J. Gullingsrud, P. Grayson, K. Schulten, A system for interactive molecular dynamics simulation, in: J.F. Hughes, C.H. Séquin (Eds.), 2001 ACM Symposium on Interactive 3D Graphics, ACM SIGGRAPH, 2001, pp. 191–194.