This article presents theoretical data on geometric and energetic features of halobenzenes and xylenes. Data were obtained from ab initio geometry optimization and frequency calculations at HF, B3LYP, MP2 and CCSD levels of theory on 6-311++G(d,p) basis set. In total, 1504 structures of halobenzenes, three structures of xylenes and one structure of benzene were generated and processed by custom-made codes in Mathematica. The quantum chemical calculation was completed in Q-Chem software package. Geometric and energetic data of the compounds are presented in this paper as supplementary tables. Raw output files as well as codes and scripts associated with production and extraction of data are also provided.
Article Title
Geometric and energetic data from quantum chemical calculations of halobenzenes and xylenes

Authors
Sopanant Datta, Taweetham Limpanuparb*

Affiliations
Science Division, Mahidol University International College, Mahidol University
999 Phutthamonthon 4, Salaya, Phutthamonthon, Nakhon Pathom 73170 Thailand

Corresponding author(s)
Taweetham Limpanuparb (taweetham.lim@mahidol.edu)

Abstract
This article presents theoretical data on geometric and energetic features of halobenzenes and xylenes. Data were obtained from ab initio geometry optimization and frequency calculations at HF, B3LYP, MP2 and CCSD levels of theory on 6-311++G(d,p) basis set. In total, 1504 structures of halobenzenes, three structures of xylenes and one structure of benzene were generated and processed by custom-made codes in Mathematica. The quantum chemical calculation was completed in Q-Chem software package. Geometric and energetic data of the compounds are presented in this paper as supplementary tables. Raw output files as well as codes and scripts associated with production and extraction of data are also provided.

Keywords
halobenzene; xylene; relative stability; steric effect

Specifications Table

| Subject          | Chemistry                                           |
|------------------|-----------------------------------------------------|
| Specific subject area | Physical and Theoretical Chemistry/Spectroscopy |
| Type of data     | Tables and Q-Chem output files                     |
| How data were acquired | Quantum chemical computation                      |
| Data format      | Raw and analysed                                    |
| Parameters for data collection | HF/6-311++G(d,p), B3LYP/6-311++G(d,p), MP2/6-311++G(d,p) and CCSD/6-311++G(d,p) |
| Description of data collection | Q-Chem 5.2.1, Developer Version                  |
| Data source location | Mahidol University, Salaya, Thailand |
|                   | Latitude and longitude: 13.792790, 100.325707     |
| Data accessibility| With the article                                    |
Value of the Data

- Relative trends and stabilities with respect to o- (ortho or 1,2-), m- (meta or 1,3-) and p- (para or 1,4-) positions have long been discussed in the literature. [1-7] The simplest model compounds for the purpose of this study are halobenzenes and xylenes. Limited experimental evidence suggest that m-xylene is more stable than its o- and p- isomers. [8]
- All 1,505 possible halobenzenes and three xylenes are explicitly shown in this paper with numbering, IUPAC name, PubChem CID and SMILES. These can be used as a reference for both theoretical and experimental work involving this class of compounds.
- Geometric and energetic data can be used for further analysis to understand relative stability of isomers. In particular, the unexpected trend in relative stability of isomers are of particular interest to scientists in a similar manner to cis and gauche effect. The data set includes many examples where steric hindrance alone fails to account for the behaviour observed in halobenzenes and xylenes. As a result, in addition to steric effect, treatment of electronic effects by quantum chemical methods are required to model these compounds. [9]
- Raw data as well as associated scripts and codes are provided so that interested researchers can reproduce our results and perform calculation at other levels of theory or for other relevant class of compounds. Vibrational spectrum and other detailed information can be extracted from output files as needed. There are many potential uses of the spectral information, for example, detection of xylene for food safety application [10] and understanding formation of polychlorinated biphenyls (PCBs). [9] The data can also be a test set for molecular modelling software packages.

Data Description

The following summary table files (.csv) are provided per level of theory (HF, B3LYP, MP2 and CCSD).

- Geometric data of 12 bond lengths, 12 bond angles and 12 torsional angles in a single csv file
- Energetic data, in separate files, include electronic energy ($E_{\text{elec}}$) in Hartree, thermal correction to enthalpy ($H_{\text{corr}}$) in kcal mol$^{-1}$, zero-point vibrational energy ($E_{\text{ZPE}}$) in kcal mol$^{-1}$ and entropy ($S$) in cal mol$^{-1}$ K$^{-1}$.

The following associated files are also provided.

- Raw Q-Chem output files (.out) for all compounds.
- Geometry in Z-matrix and Cartesian coordinate format (.xyz) for all compounds.
- Wolfram Mathematica notebook (benzene.nb) and associated script (script.txt).

Experimental Design, Materials, and Methods

A subset of halobenzenes, xylenes and related systems have been previously studied by computational methods. [3, 6, 11, 12] Our experimental design is unique for its exhaustive coverage of compound and computational methods. The four computational methods have also been previously used in our system of interest and other similar systems. MP2 may give poor result [11] due to underbinding [13] while CCSD usually overbind the van der Waals interaction. [14] These two theory can therefore provide upper and lower bound values for CCSD(T) results. [14]

A total of 1,505 unique compounds of benzene, including all degrees of substitution with F, Cl, Br, and I atoms, and three isomers of xylene were investigated. Classification and counting of the 1,505 compounds are exhaustively shown in Tables 1 and 2 with specific examples in Figures 1, 2 and 3. Table 3 summarizes the total number of Q-Chem 5.2.1 [15] output files for different class of compounds, type of calculation (geometry optimization/frequency calculation) and levels of theory (HF, B3LYP, MP2, and CCSD on 6-311++G(d,p) basis set). The output files were processed by custom-made scripts and Wolfram Mathematica 12.0 [16] codes to extract geometric and energetic data of all halobenzene compounds in a similar manner to our previous work. [17] Data from the three xylene compounds are provided for reference purpose and were read from IQmol 2.13 manually. [18]
Table 1 List of all compounds by the number of elements bonded to carbon atoms
(In total, there are 1,505 halobenzene compounds with 210 possible empirical formulas.)

| Number of elements | Distribution of elements | Number of empirical formulas | Position of elements | Number of isomers per formula | Number of structures |
|--------------------|--------------------------|-----------------------------|----------------------|-------------------------------|---------------------|
| 1                  | C₆α₆ (6)                 | (5₁)² = 5                  | n/a                  | 1                             | 5                   |
| 2                  | C₆α₅β (1-5)              | (5₁)² (2²) = 20            | 1-                   | 1                             | 20                  |
|                    | C₆α₂β₄ (2-4)             | (5₁)² (2²) = 20            | 1,2-                 | 1                             | 20                  |
|                    |                          |                            | 1,3-                 | 1                             | 20                  |
|                    |                          |                            | 1,4-                 | 1                             | 20                  |
|                    | C₆α₃β₃ (3-3)             | (5₂)² = 10                 | 1,2,3-               | 1                             | 10                  |
|                    |                          |                            | 1,2,4-               | 1                             | 10                  |
|                    |                          |                            | 1,3,5-               | 1                             | 10                  |
| 3                  | C₆αβγ₄ (1-1-4)           | (5₁)³ (3₁) = 30            | 1,2-                 | 1                             | 30                  |
|                    |                          |                            | 1,3-                 | 1                             | 30                  |
|                    |                          |                            | 1,4-                 | 1                             | 30                  |
|                    | C₆αβ₂γ₃ (1-2-3)          | (5₁)³ (3₁)² = 60           | 1,2,3-               | 2                             | 120                 |
|                    |                          |                            | 1,2,4-               | 3                             | 180                 |
|                    |                          |                            | 1,3,5-               | 1                             | 60                  |
|                    | C₆α₂β₂γ₂ (2-2-2)         | (5₃)² = 10                 | 1,2,3,4-             | 4                             | 40                  |
|                    |                          |                            | 1,2,3,5-             | 4                             | 40                  |
|                    |                          |                            | 1,2,4,5-             | 3                             | 30                  |
| 4                  | C₆αβγδ₃ (1-1-1-3)        | (5₁)⁴ (4₁) = 20            | 1,2,3-               | 3                             | 60                  |
|                    |                          |                            | 1,2,4-               | 6                             | 120                 |
|                    |                          |                            | 1,3,5-               | 1                             | 20                  |
|                    | C₆αβγ₂δ₂ (1-1-2-2)       | (5₁)⁴ (2²) = 30            | 1,2,3,4-             | 6                             | 180                 |
|                    |                          |                            | 1,2,3,5-             | 6                             | 180                 |
|                    |                          |                            | 1,2,4,5-             | 4                             | 120                 |
| 5                  | C₆αβγδε₂ (1-1-1-1-2)     | (5₁)⁵ (5₁) = 5             | 1,2,3,4-             | 12a                           | 60                  |
|                    |                          |                            | 1,2,3,5-             | 12                            | 60                  |
|                    |                          |                            | 1,2,4,5-             | 6                             | 30                  |

*aSee explicit listing of all possible isomeric structures in Figure 1.

Figure 1. List of halobenzenes with the formula C₆αβγδε where permutation of β, γ, δ and ε at four adjacent positions (1,2,3,4-) leads to 4! / 2 = 12 possible structures.
| Group of compound | Number of halogen substituents | Number of empirical formulas | Number of isomers per formula | Number of structures |
|-------------------|--------------------------------|----------------------------|----------------------------|---------------------|
| Benzene           | 0                              | $\binom{4}{0} = 1$         | n/a                        | 1                   |
| Monohalobenzene   | 1                              | $\binom{4}{1} = 4$         | 1                          | 1                   |
|                   |                                |                            |                            |                     |
|                    | 1                              | $\binom{4}{1} = 4$         |                            |                     |
|                   |                                |                            |                            |                     |
|                    | 1                              | $\binom{4}{1} = 4$         |                            |                     |
| Dihalobenzene     | 1                              | $\binom{4}{2} = 6$         |                            |                     |
|                   |                                |                            |                            |                     |
| Trihalobenzene    | 1                              | $\binom{4}{2} = 12$        |                            |                     |
|                   |                                |                            |                            |                     |
|                   | 1                              | $\binom{4}{2} = 12$        |                            |                     |
| Tetrahalobenzene  | 1                              | $\binom{4}{2} = 12$        |                            |                     |
|                   |                                |                            |                            |                     |
|                   | 2                              | $\binom{4}{2} = 12$        |                            |                     |
| Pentahalobenzene  | 1                              | $\binom{4}{1} = 4$         |                            |                     |
|                   |                                |                            |                            |                     |
|                   | 2                              | $\binom{4}{2} = 12$        |                            |                     |
See explicit listing of all possible isomeric structures in Figures 2 and 3.

Figure 2: Possible structures of pentahalobenzene with 4 different halogens acting as substituents (distribution of elements: 2-1-1-1). Structures are divided into three groups with 12, 12 and 6 structures for α atoms (any halogen listed but not H) in ortho-, meta-, and para- positions, respectively. A full list of structures of the ortho group is shown in Figure 1. Letters β, γ, δ and ε represent three different halogen atoms and a hydrogen atom.
Figure 3: List of 16 structures of hexahalobenzene with distribution of elements 2-2-1-1. Structures are organised into groups by which from left to right, the four substituents are in positions 1,2,3,4-, 1,2,3,5- and 1,2,4,5-, respectively. If switching the letters in red of a structure results in a different isomer, then that single depiction represents two different structures as shown with the notation “×2” and “×1” if otherwise. Letters β, γ, δ and ε represent different substituents of F, Cl, Br and I. Letter α represents a hydrogen atom and is omitted for simplicity.

Table 3: Summary of studied compounds, levels of theory (HF, B3LYP, MP2, and CCSD) on 6-311++G(d,p) basis set and type of calculation, (opt for geometry optimization and freq for frequency calculation)

| Group of compounds | Number of tuples | Number of structures | HF opt | HF freq | B3LYP opt | B3LYP freq | MP2 opt | MP2 freq | CCSD opt | CCSD freq |
|--------------------|------------------|----------------------|--------|---------|-----------|------------|--------|---------|---------|-----------|
| Benzene            | 1                | 1                    | all    | all     | all       | all        | all    | -       | all     | -         |
| Monohalobenzene    | 24               | 4                    | all    | all     | all       | all        | all    | -       | all     | -         |
| Dihalobenzene      | 240              | 30                   | all    | all     | all       | all        | all    | -       | all     | -         |
| Trihalobenzene     | 1,280            | 124                  | all    | all     | all       | all        | all    | -       | all     | -         |
| Tetrahalobenzene   | 3,840            | 372                  | all    | all     | all       | all        | all    | -       | all     | -         |
| Pentahalobenzene   | 6,144            | 544                  | all    | all     | all       | all        | all    | -       | all     | -         |
| Hexahalobenzene    | 4,096            | 430                  | all    | all     | all       | all        | all    | -       | all     | -         |
| Xylene             | 15               | 3                    | all    | all     | all       | all        | all    | -       | all     | -         |
| Total              | 15,640           | 1,508                | 1,508  | 1,508   | 1,508     | 1,508      | 1,508  | 1,508   | -       | 34        |

Acknowledgments
We are grateful for materials and software previously purchased by MUIC and IPST grants.
Competing Interests
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] M. El-Hamdi, O. El Bakouri El Farri, P. Salvador, B.A. Abdelouahid, M.S. El Begrani, J. Poater, M. Solà, Analysis of the Relative Stabilities of Ortho, Meta, and Para MCy(XC4H4)(PH3)2 Heterometallabenzenes (M = Rh, Ir; X = N, P; Y = Cl and M = Ru, Os; X = N, P; Y = CO), Organometallics 32 (2013) 4892-4903.
[2] J. Gorecka-Kobyłinska, M. Schlosser, Relative Basicities of ortho-, meta-, and para-Substituted Aryllithiums, The Journal of Organic Chemistry 74 (2009) 222-229.
[3] A. Kepceoğlu, Y. Gündoğdu, Ö. Dereli, H. Kiliç, Molecular Structure and TD-DFT Study of the Xylene Isomers, Gazi University Journal of Science 32 (2019) 300-308.
[4] E. Kraka, D. Cremer, Ortho-, meta-, and para-benzyne. A comparative CCSD (T) investigation, Chemical Physics Letters 216 (1993) 333-340.
[5] E. Kraka, D. Cremer, CCSD(T) Investigation of the Bergman Cyclization of Enediyne. Relative Stability of o-, m-, and p-Didehydrobenzene, Journal of the American Chemical Society 116 (1994) 4929-4936.
[6] Z. Rong, H.G. Kjaergaard, Internal Methyl Rotation in the CH Stretching Overtone Spectra of ortho-, meta-, and para-Xylene, The Journal of Physical Chemistry A 106 (2002) 6242-6253.
[7] S.K. Shin, Relative stabilities of ortho-, meta- and para-tolyl cations, Chemical Physics Letters 280 (1997) 260-265.
[8] J.D. Hepworth, D.R. Waring, M.J. Waring, M.J. Waring, Aromatic chemistry, Royal Society of Chemistry 2002, pp. 40.
[9] J. Cioslowski, G. Liu, D. Moncrieff, Energetics of the Homolytic C−H and C−Cl Bond Cleavages in Polychlorobenzenes: The Role of Electronic and Steric Effects, The Journal of Physical Chemistry A 101 (1997) 957-960.
[10] J. Shi, X. Miao, Y. Liu, Y. Tan, M. Zhang, H. Cai, Raman spectrum calculation and analysis of p-xylene, 2014 International Conference on Manipulation, Manufacturing and Measurement on the Nanoscale (3M-NANO), 2014, pp. 295-298.
[11] E. Taskinen, Relative Stabilities and Molecular Structures of Isomeric Dihalobenzenes. A DFT Study, Structural Chemistry - STRUCT CHEM 11 (2000) 293-301.
[12] G. Venkatesh, M. Govindaraju, C. Kamal, P. Vennila, S. Kaya, Structural, electronic and optical properties of 2,5-dichloro-p-xylene: experimental and theoretical calculations using DFT method, RSC Advances 7 (2017) 1401-1412.
[13] R.T. McGibbon, A.G. Taube, A.G. Donchev, K. Siva, F. Hernández, C. Hargus, K.-H. Law, J.L. Klepeis, D.E. Shaw, Improving the accuracy of Møller-Plesset perturbation theory with neural networks, The Journal of Chemical Physics 147 (2017) 161725.
[14] T.P. Tauer, M.E. Derrick, C.D. Sherrill, Estimates of the ab initio Limit for Sulfur−π Interactions: The H2S–Benzene Dimer, The Journal of Physical Chemistry A 109 (2005) 191-196.
[15] Y. Shao, Z. Gan, E. Epifanovsky, A.T. Gilbert, M. Wormit, J. Kussmann, A.W. Lange, A. Behn, J. Deng, X. Feng, Advances in molecular quantum chemistry contained in the Q-Chem 4 program package, Molecular Physics 113 (2015) 184-215.
[16] Wolfram Research Inc, Mathematica, Champaign, Illinois, 2019.
[17] K. Chinsukserm, W. Lorpaiboon, P. Teenarinamit, T. Limpanuparb, Geometric and energetic data from ab initio calculations of haloethene, haloamine, halomethylene phosphine, haloiminophosphine, halodiazene, halodiphosphene and halocyclopropane, Data in Brief 27 (2019) 104738.
[18] A.T.B. Gilbert, IQmol, http://iqmol.org, 2019.
| Other files                  | view on ChemRxiv • download file |
|-----------------------------|----------------------------------|
| codes.rar (89.09 KiB)       |                                  |
| halobenzene.rar (95.13 MiB) |                                  |
| xylene.rar (729.01 KiB)     |                                  |