





 Cite this: *RSC Adv.*, 2022, 12, 21406

 Received 2nd June 2022  
 Accepted 23rd June 2022

DOI: 10.1039/d2ra03431j

[rsc.li/rsc-advances](https://rsc.li/rsc-advances)

# Structure of benzene from mass-correlated rotational Raman spectroscopy†

 In Heo,  ‡ Jong Chan Lee,  ‡ Begüm Rukiye Özer  and Thomas Schultz \*

We present high resolution rotational Raman spectra and derived geometry parameters for benzene. Rotational Raman spectra with sub-5 MHz resolution were obtained *via* high-resolution mass-correlated rotational alignment spectroscopy. Isotopologue spectra for C<sub>6</sub>H<sub>6</sub>, <sup>13</sup>C–C<sub>5</sub>H<sub>6</sub>, C<sub>6</sub>D<sub>6</sub>, and <sup>13</sup>C–C<sub>5</sub>D<sub>6</sub> were distinguished through their correlated mass information. Spectra for <sup>13</sup>C<sub>6</sub>H<sub>6</sub> were obtained with lower resolution. Equilibrium and effective bond lengths were estimated from measured inertial moments, based on explicit assumptions and approximations. We discuss the origin of significant bias in previously published geometry parameters and the possibility to derive H,D isotope-specific bond lengths from purely experimental data.

## 1 Introduction

Rotationally resolved spectroscopy can readily resolve rotational constants and corresponding molecular inertial moments (MIM) with relative uncertainties in the 10<sup>−6</sup> range. The MIM are directly related to atomic positions within the molecule and allow to assign high-resolution molecular structures. For larger molecules, this analysis requires rotational spectra for multiple molecular isotopologues, pinpointing the position of substituted atoms in the molecular frame. We recently developed mass-correlated rotational alignment spectroscopy (CRASY), a method for the mass-selective measurement of rotational Raman spectra with broad spectral range and MHz resolution.<sup>1,2</sup> The resolution represents an order-of-magnitude improvement over that of preceding rotational Raman or Fourier-transform IR spectroscopy (FTIR)<sup>3,4</sup> and the mass-correlation facilitates the analysis of multiple isotopologue spectra in a single data-set. Here we present rotational spectra for five benzene isotopologues, estimate effective and equilibrium bond lengths, discuss the possibility to determine isotope-specific bond lengths, and explain why preceding literature values were biased.

A straightforward derivation of effective ( $r_0$ ) molecular structure from measured MIM is possible under the assumption that isotopic substitution does not affect the molecular geometry<sup>5,6</sup> (isotopic invariance approximation, IIA). Equilibrium geometries can be determined if isotope-specific rovibrational corrections are included in the analysis.<sup>7–14</sup>

Kunishige *et al.* recently used an interesting approach to directly estimate isotope-dependent  $r_0$  bond lengths in benzene from deuterated benzene MIM.<sup>15</sup> Because the isotope effect for substituting hydrogen with deuterium is much larger than that for substituting <sup>12</sup>C with <sup>13</sup>C carbon, Kunishige first used the IIA approximation to estimate the C–C bond lengths to good approximation. He then separately refined hydrogen and deuterium bond lengths using data from a series of deuterated isotopologue measurements. The authors concluded that the C–H and C–D bonds in benzene were identical to within less than one mÅ, based on a fortuitous canceling of vibrational isotope effects associated with C–H/D stretch and bend modes. A model calculation by Hirano *et al.* supported this explanation.<sup>16</sup>

The absence of a structural H/D isotope effect in benzene stands in marked contrast to the 3 mÅ to 5 mÅ bond contraction observed in small deuterated molecules<sup>9</sup> and a widely accepted “Laurie-type correction” of similar magnitude in rotational structure analysis.<sup>13,14</sup> We must therefore raise the question whether (a) benzene (and other hydrocarbon aromatic molecules) represent a special class of molecules with unexpected rovibrational properties; (b) H,D isotope effects were generally misinterpreted in the past, or (c) the data presented by Kunishige and Hirano is compatible with the traditionally expected H,D isotope effect.

Benzene has sixfold rotational symmetry ( $D_{6h}$ -symmetry) and two geometry parameters,  $r_{C-C}$  and  $r_{C-H}$ , are sufficient to describe its structure. Benzene has been thoroughly investigated with spectroscopy and *ab initio* theory and here we only give a brief summary of the relevant literature. Raman, rovibronic, and rovibrational measurements were used to characterize the rotational spectrum of benzene.<sup>2,17–38</sup> Unresolved K-splitting may explain the surprising divergence of rotational constants reported in the literature.<sup>2</sup>

UNIST (Ulsan National Institute of Science and Technology), Advanced Materials Research Building 103-413, 50 UNIST-gil, Eonyang-eup, Ulju-gun, Ulsan, 44919, South Korea. E-mail: [schultz@unist.ac.kr](mailto:schultz@unist.ac.kr); Tel: +82-52-217-5455

† Electronic supplementary information (ESI) available. See <https://doi.org/10.1039/d2ra03431j>

‡ These authors contributed equally to this work.



FTIR was used to characterize fully deuterated benzene  $C_6D_6$ ,<sup>28,39,40</sup> the  $^{13}C_6H_6$  isotopologue,<sup>28</sup> the  $^{13}C-C_5H_6$  isotopologue at natural abundance,<sup>31</sup> and the  $^{13}C_6D_6$  isotopologue.<sup>41</sup> Asymmetrically deuterated isotopologues have a small permanent dipole moment and were studied by high-resolution Fourier-transform microwave spectroscopy (FTMW).<sup>15,42,43</sup>

### 1.1 Structure determination from isotopologue inertial moments

Measured rotational constants  $A, B, C$  and the corresponding MIM  $I_{a,b,c}$  for a molecule with  $N$  atoms are directly related to the squared distance  $\langle r_{\alpha,i}^2 \rangle$  of each atomic mass  $m_i$  to the  $\alpha$  rotation axis:

$$A, B, C = \frac{\hbar^2}{2I_{a,b,c}}, \quad \text{with } I_{\alpha} = \sum_{i=1}^N m_i \langle r_{\alpha,i}^2 \rangle \quad (1)$$

Uncertainties for stable isotope masses<sup>44</sup> are in the range of  $\Delta m/m = 10^{-10}$  and can be neglected in our analysis. For the trivial example of a diatomic molecule, a single measured MIM  $I_0$  is sufficient to solve eqn (1) for an effective molecular bond length  $r_0$ .

More information is required to solve the structure of larger molecules: generally, each atomic position must be specified within a three-dimensional molecular reference frame. For an asymmetric molecule,  $(N-6)$  structural parameters are required to describe all atomic positions relative to the center-of-mass (COM). Only three MIM can be measured for a molecule (one for each spatial axis) and we can only formulate three eqn (1) and solve for three structural parameters. This is insufficient for *de novo* structure analysis, except in trivial cases. Symmetry can reduce the number of required structural parameters<sup>§</sup> but concurrently reduces the number of independently measurable MIM.

Within the IIA, *i.e.*, the assumption of  $r_X = r_Y$  for substituting isotope X with isotope Y, we can explicitly solve eqn (1) for the  $r_0$  position of an isotopically substituted atom. The complete molecular structure can be solved if a sufficient number of isotopologue MIM are determined. Multiple isotopic substitutions of the same (or a symmetry-equivalent) atom does not add new information and will not increase the number of accessible structure parameters, but can help to confirm the symmetry of the molecule and to establish confidence limits on fitted parameters.

The investigation of a complete set of isotopic substitutions (unsubstituted molecule + substitution of each symmetry-unique atom) allows us to formulate more eqn (1) than required to solve the  $r_0$  molecular structure and we can introduce an additional structural parameter, *e.g.*, an isotope-dependent bond length. This is only useful if we can identify one structural parameter that removes a significant error in the IIA approximation. In benzene, effects of zero-point vibration are almost entirely localized on the C–H bond stretch<sup>46</sup> and, following the lead of Kunishige *et al.*, we can attempt to extract isotope-dependent C–H and C–D bond lengths under the assumption that the heavier carbon atoms show negligible isotope effects. Note that, except in trivial cases, the

investigation of additional isotopologues can never yield enough information to characterize isotope-dependent bond lengths for all atoms, because any information gained through additional isotopic substitution will introduce a corresponding number of additional isotope-dependent geometry parameters.

The IIA approximation ignores that the effective bond lengths are isotope-dependent due to rovibrational coupling. This introduces a significant uncertainty in experimentally determined structures, *e.g.*, up to 10 mÅ difference in the  $r_0$  bond lengths calculated for OCS (see chapter 13.7 in ref. 45). Kraitchman proposed to calculate substitution structure parameters  $r_s$  based on the MIM differences  $I_{\alpha,iso1}^0 - I_{\alpha,iso2}^0$  between pairs of isotopologues:<sup>6,7,45</sup> the rovibrational coupling terms cancel in the MIM differences if they are similar for all observed isotopologues. The resulting  $r_s$  parameters should become isotopologue-independent and fall between  $r_0$  values and the equilibrium bond lengths  $r_e$ . Costain established a straightforward estimate for the bond length uncertainty  $\sigma_{rs} = 0.0015 \text{ \AA}/r_s$  that separates  $r_s$  parameters from  $r_e$  and  $r_0$  values.<sup>7,47</sup>

Instead of solving discrete equations for individual  $r_0$  or  $r_s$  atomic positions, iterative optimization procedures allow to determine multiple atomic positions simultaneously.<sup>48–50</sup> If the fit is over-determined (*i.e.*, the number of isotopologue MIM exceed the minimum required to solve the molecular structure), the fit approach allows to analyze parameter confidence ranges and to identify outliers in the experimental data.

Watson introduced mass-dependent rovibrational correction terms<sup>40,43</sup> to obtain  $r_m^{(1)}$  (first order approximation<sup>¶</sup>) and  $r_m^{(2)}$  (second-order approximation) geometry parameters. As presented in eqn (2), the  $r_m$  analysis requires the fitting of calculated MIM  $I_{\alpha,m}$  to measured values  $I_{\alpha,0}$ , including rovibrational correction parameters  $c_{\alpha}$  (first order correction) and  $d_{\alpha}$  (second order correction) for each rotational axis. The correction terms represent the sum of rovibrational effects with respect to rotation around the corresponding axis  $\alpha$ . Resulting  $r_m$  geometry parameters approximate corresponding equilibrium values  $r_e$ .

$$I_{0,\alpha} = I_{m,\alpha} + c_{\alpha} \cdot \sqrt{I_{m,\alpha}} + d_{\alpha} \cdot \left( \frac{\prod_{i=1}^N m_i}{\sum_{i=1}^N m_i} \right)^{1/(2N-2)} \quad (2)$$

The orientation of principal rotational axes within the molecular frame can differ between isotopologues. This can be rectified within the first-order approximation by including off-diagonal correction terms  $c_{\alpha\beta}(\sqrt{I_m})_{\alpha\beta}$ , which project the corrections into an invariant molecular coordinate frame. Resulting geometry parameters are labeled  $r_m^{(1r)}$  and  $r_m^{(2r)}$ . Fitting of  $r_m$  geometry parameters represents the current state-of-the-art for estimating equilibrium bond lengths from

¶ The  $r_m^{(1)}$  terms are based on the approximate mass-dependence expected for harmonic vibration.

||  $(\sqrt{I_m})_{\alpha\beta}$  ( $\alpha, \beta \in a, b, c$ ) represent the tensor elements created by rotating the square root of the calculated principal inertial moments vector  $\sqrt{I_{m,\alpha}}$  into the reference coordinate system. Off-diagonal parameters  $c_{\alpha\beta}$  are treated as independent fit parameters.

§ The number of structural parameters required to characterize a particular molecule is equal to the number of totally symmetric vibrations.<sup>45</sup>



isotopologue rotational constants, was found to reproduce synthetic data with high accuracy, and was used to obtain meaningful bond parameters from experimental data.<sup>13,14,51</sup>

The large and localized effects of hydrogen-atom substitution can be separately addressed by a Laurie-type correction,<sup>9,13</sup> substituting the bond length for each X-H bond length (H = H, D, T) with the mass-corrected value in eqn (3). The Laurie coefficient was found to have a fairly universal value of  $\delta_{\text{H}} \sim 10 \text{ \AA} \text{ u}^{1/2}$ . Laurie-corrected fit methods and geometry parameters are marked with a superscript (L).

$$r^{\text{L}}(\text{XH}) = r(\text{XH}) + \delta_{\text{H}} \left[ \frac{\sum_{i=1}^N m_i}{m_{\text{H}} \left( \sum_{i=1}^N m_i - m_{\text{H}} \right)} \right]^{1/2} \quad (3)$$

## 2 Experimental part

The experimental technique of correlated rotational alignment spectroscopy is described in detail in our previous work.<sup>1,52</sup> Briefly, a pump pulse (800 nm, 100  $\mu\text{J}$ , 1 ps) from a 1 kHz amplified femtosecond Ti:Sa laser excited a coherent superposition of rotational states *via* rotational Raman excitation, creating a rotational wave packet. A time-delayed probe pulse (200 nm, 1  $\mu\text{J}$ , 45 fs) from a second 1 kHz Ti:Sa amplifier probed the evolution of the rotational wave packet *via* resonance enhanced two-photon ionization. Resulting ions were detected in a Wiley McLaren time-of-flight mass spectrometer. Interference between the coherent rotational quantum states modulated the ionization efficiency, leading to delay dependent signal modulation in the ion signals. The rotational Raman frequencies encoded in the signal modulation were analyzed by Fourier-transformation of the delay dependent ion signals.

Helium carrier gas flowed through a sample holder containing a 50 : 50 mixture of benzene (Sigma Aldrich, HPLC grade) and D<sub>6</sub>-benzene (Sigma Aldrich, 99% deuterated) or benzene and <sup>13</sup>C<sub>6</sub>-benzene (Sigma Aldrich, 99 atom%). The benzene-helium mixture, at 5 bar pressure, was expanded through a pulsed valve (Even Lavie, E.L-7-4-2007-HRR), running at 500 Hz. The resulting cold molecular beam passed through a skimmer and reached the spectrometer region, where it interacted with pump and probe laser pulses. The time delay between pump and probe pulses was controlled by a combination of an opto-mechanical delay line (folded length of 4.8 m) and an electronic pulse-selection delay. The electronic pulse selection added discrete delays in multiples of 12.5 ns, corresponding to the pulse-to-pulse delay in the Ti:Sa laser oscillator. The oscillator repetition rate was measured against a GPS-stabilized clock to calibrate the time delay axis with relative uncertainties  $\Delta t/t \ll 10^{-9}$ . Resulting rotational Raman spectra can therefore be considered as absolute frequency spectra within the statistical uncertainty. For details about the absolute frequency calibration (a time domain equivalent to frequency comb spectroscopy), we refer the reader to our earlier work.<sup>1</sup>

The data presented for C<sub>6</sub>H<sub>6</sub>, <sup>13</sup>C-C<sub>5</sub>H<sub>6</sub>, C<sub>6</sub>D<sub>6</sub>, and <sup>13</sup>C-C<sub>5</sub>D<sub>6</sub> was obtained by sampling 20 000 mass spectra over a delay

range of 200 ns. Delays were sampled in random multiples of 1 ps (10% random sparse sampling). The data presented for <sup>13</sup>C<sub>6</sub>H<sub>6</sub> was obtained by sampling 20 153 mass spectra over a delay range of 20.153 ns, with constant 1 ps step size. Rotational Raman spectra were obtained by Fourier-transformation of mass-selected, delay-dependent signal amplitudes after suitable baseline corrections. Resulting rotational Raman spectra had a 500 GHz spectral range and a non-apodized full-width-at-half-maximum (FWHM) resolution limit<sup>3</sup> of 3.1 MHz (30 MHz) for the first (second) sample. A break-down of the laser system kept us from measuring higher resolution spectra for the second sample. In the latter case, the delay range was insufficient for calibration against the external clock and we calibrated residual CS<sub>2</sub> signals against literature values.

## 3 Results

### 3.1 Spectroscopic results

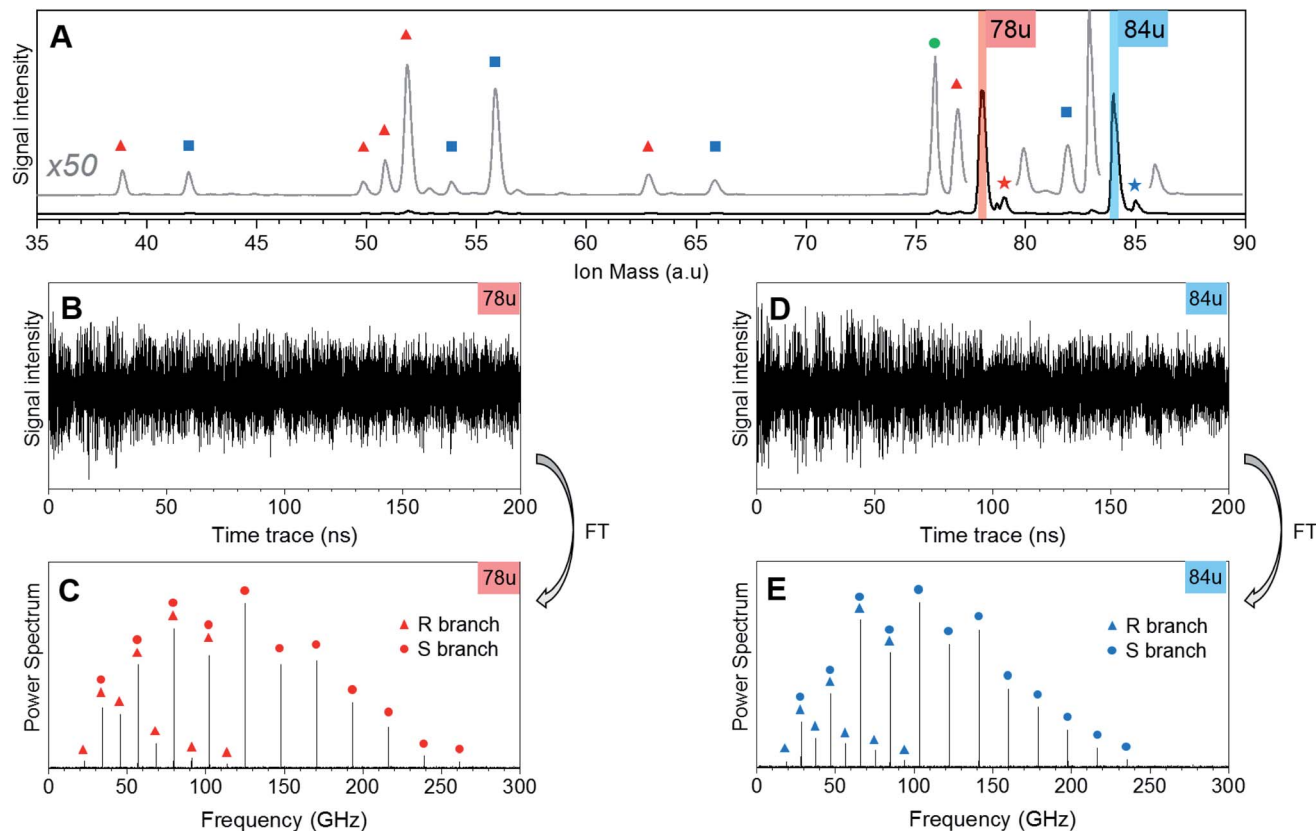
Fig. 1 shows CRASY experimental data for the first sample. Results from other data sets with different sampling of the delay axis gave comparable results. The integrated mass spectrum in Fig. 1A shows parent ions and fragments for C<sub>6</sub>H<sub>6</sub> and C<sub>6</sub>D<sub>6</sub>, as well as weak signals for naturally occurring <sup>13</sup>C isotopologues. Molecular fragments were assigned to parent species through a comparison of their correlated rotational spectra, as described in more detail below, and mass peaks are marked accordingly. Ion signals at mass 78 u and 79 u stem from C<sub>6</sub>H<sub>6</sub> and its <sup>13</sup>C-isotopologue. Masses 84 u and 85 u stem from C<sub>6</sub>D<sub>6</sub> and its <sup>13</sup>C-isotopologue. Further mass signals in the range 80 u to 86 u were due to additional benzene isotopologues but were not strong enough to assign correlated rotational spectra. A signal at 76 u, stems from carbon disulfide, which remained in the sample line from earlier measurements.

The fragments of the undeuterated and deuterated benzene isotopologues appeared as pairs with an integer mass spacing reflecting the number of hydrogen/deuterium atoms in the fragments: 39 u and 42 u (C<sub>3</sub>H<sub>3</sub> and C<sub>3</sub>D<sub>3</sub>), 51 u and 54 u (C<sub>4</sub>H<sub>3</sub> and C<sub>4</sub>D<sub>3</sub>), 52 u and 56 u (C<sub>4</sub>H<sub>4</sub> and C<sub>4</sub>D<sub>4</sub>), 63 u and 66 u (C<sub>5</sub>H<sub>3</sub> and C<sub>5</sub>D<sub>3</sub>), 77 u and 82 u (C<sub>6</sub>H<sub>5</sub> and C<sub>6</sub>D<sub>5</sub>). The mass channel 50 u (C<sub>4</sub>H<sub>2</sub>) showed a benzene fragment but the corresponding deuterated signal at mass 52 u could not be assigned due to the dominant C<sub>4</sub>H<sub>4</sub> fragment signal at this mass.

Fig. 1B and D show the ion mass signal modulations for C<sub>6</sub>H<sub>6</sub> and C<sub>6</sub>D<sub>6</sub> as a function of the 200 ns pump-probe time delay. The time traces were Fourier-transformed to obtain their frequency-domain rotational Raman spectra as shown in Fig. 1C and E, respectively. R-branch ( $\Delta J = \pm 1$ ) and S-branch ( $\Delta J = \pm 2$ ) transitions were assigned and are marked in the spectra. No bands were observed in the higher frequency region (300–500 GHz) of the spectrum. The experimental FWHM resolution for all bands was close to 4.5 MHz.

Note that the excitation and probing of the rotational Raman wave packets occurs in the neutral ground state, before electronic excitation and ionization induces molecular fragmentation. The rotational spectra correlated to ionic fragments is therefore identical to the parent spectrum and allows an unambiguous assignment of each fragment to its parent





**Fig. 1** CRASY data for a mixed sample of  $C_6H_6$  and  $C_6D_6$ . (A) Excerpt from the mass spectrum of the heterogeneous sample. The trace in grey shows the same spectrum with enlarged ordinate and vertical offset. Fragments of  $C_6H_6$  (78 u) are marked with red triangles, fragments of  $C_6D_6$  (84 u) are marked with blue squares, signals for  $^{13}C$  isotopologues are marked with stars and a signal from carbon disulfide (76 u) is marked with a green circle. (B and D) Delay-dependent traces for the 78 u and 84 u mass channels. (C and E) Rotational Raman spectra for the 78 u and 84 u mass channels, obtained by Fourier-transformation of respective traces (B and D). Transitions were assigned to the R branch (triangles,  $J \leftrightarrow J + 1$  transitions) and S branch (circles,  $J \leftrightarrow J + 2$  transitions) as indicated.

species. Fig. 2 compares sections of the rotational Raman spectrum correlated to mass 78 u ( $C_6H_6$ ) to those of all mass peaks marked by red triangles in Fig. 1A. The spectra are identical and transition frequencies agree with deviations well below 1 MHz. All marked ion signals can therefore be assigned as fragmentation products of  $C_6H_6$ .

In Fig. 3, we compare the rotational Raman spectrum correlated with mass 84 u ( $C_6D_6$ ) to that correlated with all ion signals marked by blue squares in Fig. 1A. The spectra are identical and all corresponding ion signals can therefore be assigned as fragmentation products of  $C_6D_6$ .

Our experiment was sensitive enough to resolve naturally occurring  $^{13}C$  isotopologues ( $^{13}C$  abundance of 1.1%) for the benzene species. Fig. 4 shows rotational Raman spectra for the  $1-^{13}C$  carbon isotopologues of benzene and  $D_6$ -benzene, compared to corresponding simulations created in the PGOPHER program.<sup>53</sup>

The measurements performed on sample containing  $^{13}C_6H_6$  gave spectra with lower resolution and signal quality. Because the parent ion signals were in saturation, we analyzed the rotational spectrum correlated to the sum of fragment species at mass 42 u, 56 u, and 68 u. The relevant mass- and rotational spectra are shown in the ESI† (section 1).

Rotational transitions for all observed benzene isotopologue species were assigned by comparison with simulated spectra in the PGOPHER program.<sup>53</sup> PGOPHER was then used to fit rotational constants to the assigned lines. Resulting spectroscopic constants for symmetric and asymmetric isotopologues are listed in Tables 1 and 2. Because we only observed low rotational states in our cold molecular beam, uncertainties for fitted distortion constants were sometimes excessively large and we held corresponding values to calculated distortion constants obtained from MP2 calculations with aug-cc-pVTZ basis set (obtained using the `Freq = (vibrot)` keyword in Gaussian<sup>54</sup>). For line-lists and details on the line assignments and fits, see sections (1) and (2) in the ESI† section 1.

The rotational constant for benzene determined here was significantly higher than that reported in ref. 2. This is due to the assignment of higher rotational transitions (here up to  $J = 11 \leftrightarrow 13$ ), which requires the inclusion of higher-order distortion constants. Omitting higher  $J$  transitions reduced the fitted rotational constant and brought it closer to literature values. Because, at current resolution, the K-splitting in symmetric benzene isotopologues is not resolved, the distortion constants remain badly determined and we cannot give definitive,



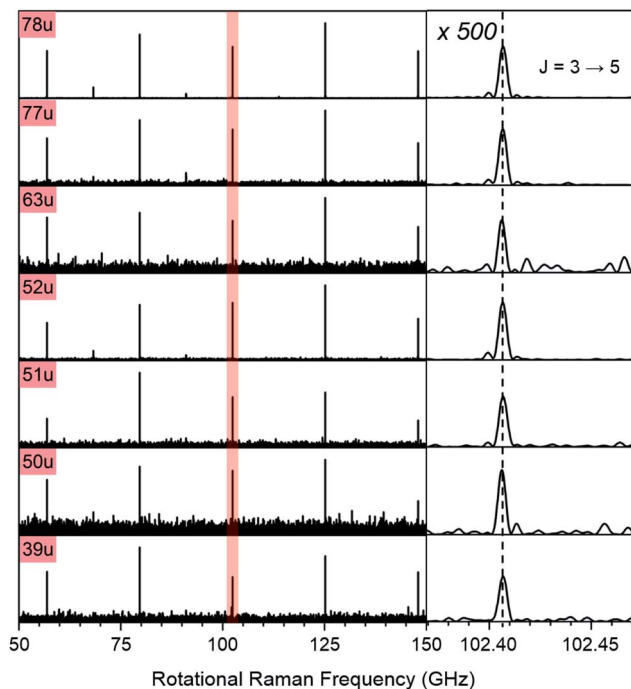


Fig. 2 Section of the rotational Raman spectra for  $C_6H_6$  (78 u) and its six cationic fragments (39 u, 50 u, 51 u, 52 u, 63 u, 77 u, marked with red triangles in Fig. 1A). The enlarged inset on the right shows the  $J = 3 \leftrightarrow 5$  transition at 102.406 GHz with 500-fold enlarged abscissa to illustrate the exact correspondence of peak positions.

'correct' rotational constants. But the achieved accuracy is sufficient for a structure analysis, as described below.

### 3.2 Structure analysis

All code used for the structure analysis is available in form of two Python scripts. The first script gives a step-by-step derivation of the relevant equations (ESI,† section 3.1) and the second script offers optimized code and a graphical user-interface for structure fitting (ESI,† section 3.2). Fig. 5 shows the structure and relevant structural parameters of benzene. For the  $D_{6h}$ -symmetric planar oblate-top molecules  $C_6H_6$  and  $C_6D_6$ , it is straightforward to write eqn (1) that relate inertial moments  $I_a = I_b = \frac{1}{2}I_c$  to bond lengths  $r_{CC}$  and  $r_{CH}$ :

$$I_b = 3m_C r_{CC}^2 + 3m_H (r_{CC} + r_{CH})^2 \quad (4)$$

$$I_b = 3m_C r_{CC}^2 + 3m_D (r_{CC} + r_{CD})^2 \quad (5)$$

Within the IIA approximation ( $r_{CH} = r_{CD}$ ), and using the inertial moments  $I_b = \hbar^2/2B$  corresponding to rotational constants in Table 1, eqn (2) and (3) can be solved analytically to obtain bond lengths  $r_{0,CH} = 1.0804(12)$  Å and  $r_{0,CC} = 1.3971(11)$  Å (ESI,† section 3.1.3). The propagated uncertainties from the measured rotational constants were extremely small and values in brackets instead give the Costain errors for the corresponding last digits,<sup>7,12</sup> which accounts for the expected uncertainty range between  $r_0$ ,  $r_s$ , and  $r_e$  parameters. We also solved the corresponding equations for the isotopologue pair  $C_6H_6$  and  $^{13}C_6H_6$  and obtained bond lengths

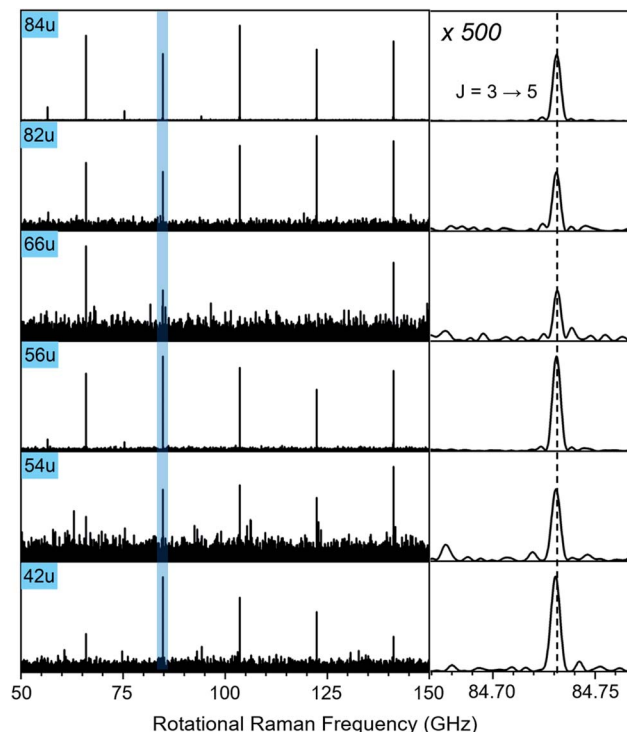


Fig. 3 Section of the rotational Raman spectra for  $C_6D_6$  (84 u) and its five cationic fragments (42 u, 54 u, 56 u, 66 u, 82 u, marked with blue squares in Fig. 1A). The enlarged inset on the right shows the  $J = 3 \leftrightarrow 5$  transition at 84.731 GHz with 500-fold enlarged abscissa to illustrate the exact correspondence of peak positions.

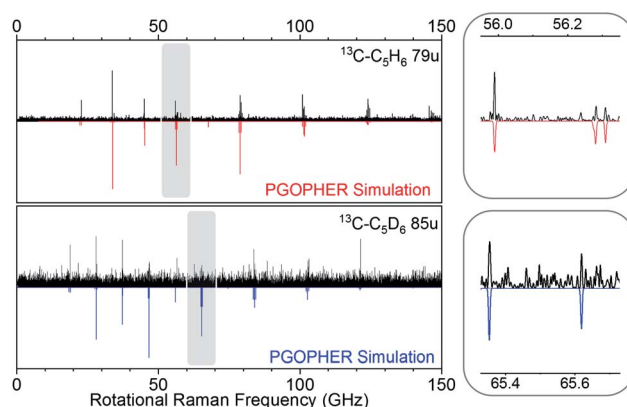


Fig. 4 Comparison of experimental simulated spectra for heavy carbon isotopologues at 79 u (top,  $^{13}C-C_5H_6$ ) and 85 u (bottom,  $^{13}C-C_5D_6$ ).

$r_{0,CH} = 1.1058(12)$  Å and  $r_{0,CC} = 1.3938(11)$  Å. The bond lengths calculated for the two isotopologue pairs differ by a value far greater than the Costain uncertainty. This illustrates the inadequacy of the IIA assumption and should not be surprising: Costain's original manuscript already stated that "equations [...] could not be satisfied unless the H and D atoms were assumed to have different effective positions."

To obtain results beyond the IIA approximation, we fitted a variety of geometric parameters to measured inertial moments



**Table 1** Ground state rotational constants (in MHz), fitted to observed transitions at mass 78 u ( $C_6H_6$ ) and 84 u ( $C_6D_6$  or  $^{13}C_6H_6$ ). Numbers in round brackets give the  $1 - \sigma$  standard deviation in the corresponding last digits

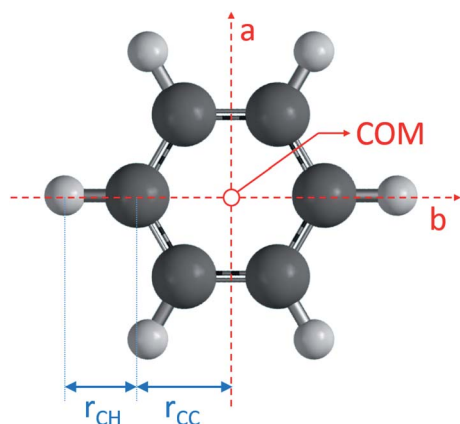
	$C_6H_6$	$C_6D_6$	$^{13}C_6H_6$
$B_0$	5689.2855(54)	4707.3175(34)	5337.884(51)
$D_J$	$0.79(19) \times 10^{-3}$	$0.64(10) \times 10^{-3}$	$0.84(74) \times 10^{-3}$
$D_{JK}$	$-0.78(51) \times 10^{-3}$	$-0.93(27) \times 10^{-3}$	$-4.1(23) \times 10^{-3}$
$D_K$	$^a[0.935 \times 10^{-3}]$	$^a[0.536 \times 10^{-3}]$	$^a[0.838 \times 10^{-3}]$
$H_J$	$-5.1(20) \times 10^{-6}$	$-1.38(92) \times 10^{-6}$	
$H_{JK}$	$-8.1(91) \times 10^{-6}$	$-5.2(41) \times 10^{-6}$	
$H_{KJ}$	$55(17) \times 10^{-6}$	$0.221(74) \times 10^{-6}$	
Lines	16	18	16

<sup>a</sup> Values from MP2 aug-cc-pVTZ calculation.

**Table 2** Ground state rotational constants (in MHz) fitted to observed transitions at mass 79 u ( $^{13}C-C_5H_6$ ) and 85 u ( $^{13}C-C_5D_6$ ). Numbers in round brackets give the  $1 - \sigma$  standard deviation in the corresponding last digits

	$^{13}C-C_5H_6$	$^{13}C-C_5D_6$
$A_0$	5689.474(18)	4707.541(36)
$B_0$	5568.473(23)	4624.188(31)
$C_0$	2868.6(73)	2332(16)
$\Delta_J$	$2.77(53) \times 10^{-3}$	$^a[0.709 \times 10^{-3}]$
$\Delta_{JK}$	$-10.3(29) \times 10^{-3}$	$^a[-1.169 \times 10^{-3}]$
$\Delta_K$	$^a[0.902 \times 10^{-3}]$	$^a[0.521 \times 10^{-3}]$
$\delta_J$	$^a[16.906 \times 10^{-6}]$	$^a[8.121 \times 10^{-6}]$
$\delta_K$	$^a[276.963 \times 10^{-6}]$	$^a[0.164 \times 10^{-3}]$
Assigned lines	35	28

<sup>a</sup> Values from MP2 aug-cc-pVTZ calculation.



**Fig. 5** Benzene molecular geometry and the  $a$ ,  $b$  principal rotational axes. The molecule has  $D_{6h}$  symmetry and is fully described by two bond lengths  $r_{CC}$  and  $r_{CH}$ . Note that  $r_{CC}$  is identical to the distance of the carbon atoms from the center-of-mass (COM).

$I_0$  (ESI,† section 3.1.5). Atomic coordinates were calculated from a minimal set of geometry parameters while assuming  $D_{6h}$  symmetry for all benzene isotopologues. Molecular moments-of-inertia tensors  $I_{\alpha\beta}$  ( $\alpha, \beta \in a, b, c$ ) were calculated from atomic masses and coordinates. The corresponding principal moments

of inertia  $I_{calc.}$  were obtained as eigenvalues of the secular equation  $I_{\alpha\beta} - I_{calc.,\alpha,\alpha}$  (ESI,† section 3.1.4, see also Chapter 2 of Gordy & Cook<sup>45</sup> or ref. 6). A Levenberg–Marquardt nonlinear fit was used to determine geometric parameters that minimized the squared residuals  $(I_0 - I_{calc.})^2$ . To account for very different uncertainties  $\sigma$  in experimental MIM, we used a weight function  $w_i = (\sigma_i^2 + b^2)^{-1}$  with  $b = 0.0005 \mu\text{Å}^2$ , as suggested by Eijck.<sup>47\*\*</sup>

We fitted the benzene structure based on our own rotational constants, as summarized in Tables 1 and 2, and also based on additional rotational constants from the literature.<sup>15,28,31,32,39–41,43,55</sup> If corresponding authors published multiple values for the same isotopologue, we only used the newest or most accurate published values. A summary of relevant literature constants is given in Table 3.

In a first approach, we fitted effective bond lengths  $r_0$  (ESI,† section 3.1.5). We implemented a two-parameter fit of  $r_{CC}$  ( $=r_{C^{13}C}$ ) and  $r_{CH}$  ( $=r_{CD}$ ), corresponding to the IIA approximation, a three parameter fit of  $r_{CC}$  ( $=r_{C^{13}C}$ ),  $r_{CH}$ , and  $r_{CD}$ , and a four parameter fit with distinct  $r_{CH}$ ,  $r_{CD}$ ,  $r_{CC}$ , and  $r_{C^{13}C}$ . Fit results and related literature values are summarized in Table 4. For  $r_0$  structure analysis, we only fitted  $a$  and  $b$  axis constants because meaningful  $c$ -axis constants were only available for deuterated isotopologues and showed large systematic deviations due to the inertial defect. A Monte-Carlo analysis was performed to evaluate the propagation of experimental uncertainties (ESI,† section 3.1.6).

The two parameter fit of our own isotopologue rotational constants (Table 4, row 1) closely reproduced the results from the analytical calculation for the  $C_6H_6$  and  $C_6D_6$  isotopologue pair and are in good agreement with literature values (Table 4, rows 12–15). Including additional literature constants in the analysis led only to slight changes in the fitted bond lengths (Table 4, row 2). The fitted  $r_0$  bond lengths changed significantly if we removed all deuterated isotopologue data from the fit, with the fit result converging to that from the analytical calculation for the  $C_6H_6$ ,  $^{13}C_6H_6$  isotopologue pair (Table 4, row 3). The latter values represent the best estimates for the actual  $r_0$  bond lengths: the isotope dependence for the C–H/D bond length is expected to be much larger than that for the C–C bond length on account of the larger amplitude vibrational motion of the light hydrogens as compared to the heavier carbons. The approximation  $r_{CH} = r_{CD}$  required for the analysis of deuterated benzenes is therefore significantly worse than the approximation  $r_{CC} = r_{C^{13}C}$  required for the analysis of  $^{13}C$ ,  $^{12}C$  isotopologues.

The three parameter fits (Table 4, rows 4 and 5) indicated a significant difference between C–H and C–D bond lengths, with  $r_{CH} - r_{CD} = 8.6(4.9)$  mÅ (our data) or  $11.5(2.8)$  mÅ (including literature data). Note that the  $r_C$  and  $r_{CH}$  bond lengths agree well with our best estimates from the 2-parameter fit (Table 4, row 3) but not with results obtained from deuterated isotopologue data (Table 4, rows 1, 2, 12–15). This confirms our assumptions about the severity of the IIA approximation for H/D versus  $^{12}C/^{13}C$  isotopologues. Our results are in clear

\*\* Eijck's fit weights account for the fact that model errors are larger than the experimental uncertainty for high resolution data: for small values of  $\sigma$ , fit weights  $(\sigma_i^2 + b^2)^{-1} \approx b^{-2}$  are constant. For low resolution data ( $\sigma_i \gg b^2$ ), fit weights scale proportional to  $\sigma_i^{-2}$ .



**Table 3** Benzene isotopologue rotational constants (in MHz) used for structure analysis. Values marked in grey were outliers and were excluded from the analysis. Number in brackets denote the estimated uncertainty ( $1 - \sigma$ ) in the corresponding last digits

Isotopologue	$A_0$	$B_0$	$C_0$	Symmetry	Author	Ref.
$C_6H_6$	—	5688.9220(060)	—	Oblate pl.	Pliva1990	28
$C_6H_6$	—	5689.2664(060)	—	Oblate pl.	Hollenstein1990	31
$C_6H_6$	—	5689.2781(010)	—	Oblate pl.	Juntilla1991	32
$C_6H_6$	—	5689.2120(009)	—	Oblate pl.	Doi2004	55
$C_6H_6$	—	5689.2670(005)	—	Oblate pl.	Lee2019	2
$C_6H_6$	—	5689.2855(005)	—	Oblate pl.	In2021	This work
$^{13}C-C_5H_6$	5689.474(018)	5568.473(023)	2868.600(730)	Planar	In2021	This work
$D_1-C_6H_5$	5689.144(006)	5323.934(006)	2749.674(006)	Planar	Oldani1984	42
$D_1-C_6H_5^a$	5689.143(006)	5323.933(006)	2749.675(006)	Planar	Kunishige2015	15
$o-D_2-C_6H_4$	5498.062(004)	5164.242(004)	2662.496(004)	Planar	Oldani1988	43
$o-D_2-C_6H_4^b$	5498.032(009)	5164.213(009)	2662.466(006)	Planar	Kunishige2015	15
$m-D_2-C_6H_4$	5502.669(007)	5152.057(006)	2660.358(006)	Planar	Oldani1988	43
$m-D_2-C_6H_4^a$	5502.667(009)	5152.053(009)	2660.352(009)	Planar	Kunishige2015	15
$o-D_3-C_6H_3^c$	5168.017(015)	5151.933(060)	2579.579(006)	Planar	Kunishige2015	15
$o-D_4-C_6H_2^c$	5163.715(006)	4846.814(006)	2499.792(003)	Planar	Kunishige2015	15
$m-D_4-C_6H_2^c$	5151.993(120)	4850.312(120)	2497.902(030)	Planar	Kunishige2015	15
$o-D_5-C_6H_1^c$	4998.170(150)	4707.221(150)	2423.972(060)	Planar	Kunishige2015	15
$C_6D_6$	—	4707.125(006)	—	Oblate pl.	Doi2004	55
$C_6D_6$	—	4707.312(052)	—	Oblate pl.	Pliva1989	39
$C_6D_6$	—	4707.327(006)	—	Oblate pl.	Snels2002	40
$C_6D_6$	—	4707.318(003)	—	Oblate pl.	In2021	This work
$^{13}C-C_5D_6$	4707.541(036)	4624.188(031)	2332(16)	Planar	In2021	This work
$^{13}C_6H_6$	—	5337.925(060)	—	Oblate pl.	Pliva1990	28
$^{13}C_6H_6$	—	5337.884(051)	—	Oblate pl.	In2021	This work
$^{13}C_6D_6$	—	4464.371(024)	—	Oblate pl.	Pliva1991	41

<sup>a</sup> Kunishige's<sup>15</sup> re-evaluated constants from Oldani1988,<sup>43</sup> with distortion constants  $D_J$ ,  $D_{JK}$ , and  $D_K$  fixed to the averaged values of  $C_6H_6$  and  $C_6D_6$ .

<sup>b</sup> Kunishige's<sup>15</sup> re-evaluated constants from Oldani1988,<sup>43</sup> including one additional transition and with distortion constants  $D_J$ ,  $D_{JK}$ , and  $D_K$  fixed to the averaged values of  $C_6H_6$  and  $C_6D_6$ . <sup>c</sup> Distortion constants  $D_J$ ,  $D_{JK}$ , and  $D_K$  fixed to the averaged values of  $C_6H_6$  and  $C_6D_6$ .

contradiction to the absence of an H,D geometric isotope effect discussed by Kunishige<sup>15</sup> and Hirano.<sup>16</sup> If we perform our 3-parameter fit with the deuterated isotopologue data-set reported by Kunishige *et al.* we reproduce their fit result but find excessive parameter uncertainties (Table 4, row 6).

In the four parameter fit, geometry parameters for  $r_{12C/^{13}C}$  and  $r_{H/D}$  were strongly correlated and fit results remained badly determined. This fit approach required an assumption about how much of the effective displacement of a  $^{13}C$  versus  $^{12}C$  atom leads to a corresponding displacement of the attached H-atom.<sup>††</sup> Treating this as an independent fit parameter gave an optimized H displacement value of 0.4 times the displacement of the bonded carbon atom, but the uncertainty of this parameter and of fitted bond lengths exceed the parameter values themselves. We conclude that a direct fit of isotope dependent  $r_0$  parameters for all atoms is not feasible. This agrees with our expectations outlined in the introduction: isotope-dependent bond lengths can not be obtained for all substituted atoms, because any

information gained through additional isotopic substitution will introduce a corresponding number of additional geometry parameters.

The difficulty to determine H,D isotope-specific bond lengths from purely deuterated benzene data-sets can be understood from a consideration of the underlying mathematical model: the fit must solve a set of equations that relate measured inertial moments to atomic positions. A solution is only possible when the number of equations is identical, or larger, than the number of unknown parameters. At first glance this should not be an issue because we have rotational constants for 8 different deuterated benzene isotopologues at our disposal. But H atoms in benzene are equivalent by symmetry and the resulting equations are linearly dependent to very high degree.<sup>‡‡</sup> The large parameter uncertainty is therefore a direct consequence of the limited data-set. With our experimental data for  $^{13}C$  substituted benzenes, we added a set of linearly independent equations and therefore obtained well-determined  $r_0$  geometry parameters for C–C, C–H and C–D bonds. We conclude that the absence of an H,D geometric isotope effect reported by Kunishige is an artefact of a highly undetermined fit.

<sup>††</sup> Note that the analysis, based on eqn (1), determines expectation values for  $\langle r^2 \rangle$  and not  $\langle r \rangle^2$ . Harmonic vibrations around  $r_e$  are symmetric and do not affect  $\langle r \rangle^2$  but increase  $\langle r^2 \rangle$ , *i.e.*, lead to an apparent isotope-dependent displacement that does not correspond to a physical displacement of the corresponding atom. Anharmonic vibrations lead to an actual mass-dependent displacement of the nuclear position ( $r_e \neq r_0$ ). Also note that we ignore the possible displacement of carbon atoms bonded to the substituted carbon, as this would break our underlying assumption of  $C_6$  rotational symmetry.

<sup>‡‡</sup> The small center-of-mass shift in asymmetrically deuterated isotopologues breaks the linear dependence to some degree.



**Table 4** Geometry parameters for benzene (for details, see text). Numbers in round brackets give the Costain error (for  $r_0$  fit results) or the fit error (for  $r_0^{(L)}$  and  $r_m$  fit results)

Fit of effective geometry parameters				
Row	Fit type	$r_C$ (Å)	$r_{CH}$ (Å)	$r_{CD}$ (Å)
1	$r_0$ , 2-parameters <sup>a</sup>	1.3971(11)	1.0806(14)	
2	$r_0$ , 2-parameters <sup>b</sup>	1.3971(11)	1.0804(14)	
3	$r_0$ , 2-parameters <sup>c</sup>	1.3938(2)	1.1059(7)	
4	$r_0$ , 3-parameters <sup>a</sup>	1.3945(12)	1.1003(45)	1.0917(28)
5	$r_0$ , 3-parameters <sup>b</sup>	1.3937(11)	1.1069(28)	1.0953(19)
6	$r_0$ , 3-parameters <sup>d</sup>	1.3972(1 410 525)	1.0804(10 881 503)	1.0804(6 149 729)
7	$r_0^{(L)b}$	1.3937(10)	1.1066(233)	1.0952(18)
Geometry fit using mass-weighted rovibrational and Laurie corrections				
Row	Fit type	$r_{CC}$ (Å)	$r_{CH}$ (Å)	$\delta_H$ (Å u <sup>1/2</sup> )
7'	$r_0^{(L)b}$	1.3937(11)	1.0671(15)	0.0393(21)
8	$r_m^{(1)e}$	1.3912(4)	1.0817(5)	
9	$r_m^{(1L)e}$	1.3939(27)	1.0660(136)	0.0421(432)
10	$r_m^{(2)e}$	1.3913(2)	1.0814(3)	
11	$r_m^{(2L)e}$	1.3921(15)	1.0769(76)	0.012(24)
Literature values				
Row	Fit type	$r_{CC}$ (Å)	$r_{CH}$ (Å)	$r_{CD}$ (Å)
12	$r_0$ , Kunishige2015	1.3971	1.0805	1.0805
13	$r_0$ , Doi2004	1.3971	1.0807	
14	$r_0$ , Pliva1990	1.3969	1.0815	
15	$r_0$ , Baba2011	1.3969	1.0817	
16	$r_c$ , Kunishige2015	1.3892	1.0864	
17	$r_c$ , Pliva1990	1.3893	1.0857	
18	$r_c$ , Pliva1991	1.3902	1.0862	
19	$r_c$ , CCSD(T) <sup>f</sup>	1.3920	1.0802	

<sup>a</sup> Based on  $A, B$  constants presented in Tables 1 and 2. <sup>b</sup> Based on  $A, B$  constants in Table 3. <sup>c</sup> Based on  $B$  constants for <sup>12</sup>C, <sup>13</sup>C isotopologues in Table 3. <sup>d</sup> Based on  $A, B, C$  constants for deuterated isotopologues in ref. 15. <sup>e</sup> Based on  $A, B, C$  constants as listed in Table 3. <sup>f</sup> Reference values from the CCCBDB database<sup>56</sup> for a coupled-cluster (full) calculation with aug-cc-pVTZ basis.

The assumption of different, but isotopologue-independent CH and CD bond lengths in the 3-parameter fits ignores that the isotopologue masses are different: not only the H/D mass affects the vibrational amplitude, but also the mass of the remainder of the molecule. The Laurie-correction (eqn (3)) properly accounts for the reduced mass of the vibrational modes. Including the Laurie correction in the fitting of effective  $r_0$  bond length gave an excessively large Laurie parameter of  $\delta_H = 39 \text{ mÅ u}^{1/2}$  (Table 4, rows 7 and 7') and correspondingly large values for  $r_{CH}$  and  $r_{CD}$ .

Results from Monte Carlo analyses (ESI,† section 3.1.6) agreed with the fit results and gave propagated uncertainties well below the fit uncertainties. This indicates that model errors dominate in our analysis. We found a strong linear correlation between fit parameters in 3 and 4 parameter fits. Fitting of synthetic data verified that this is a fundamental limitation of the fit model and is not due to inaccuracies in the available isotopologue rotational constants (ESI,† section 3.1.7).

Watson's  $r_m$  formalism<sup>13</sup> accounts for the mass-dependence of rovibrational correction terms (ESI,† sections 3.1.8 and 3.2). This fit approach models the magnitude of rovibrational corrections

based on our physical understanding (within first- and second-order approximations) of rovibrational effects and allows us to estimate equilibrium bond lengths. To account for the symmetry of benzene, we removed redundant in-plane rovibrational correction terms. Some fit uncertainty arose from the choice of axes along which the rovibrational correction terms were applied. § We

§§ In the literature, first-order ( $r_m^{(1)}$ ) rovibrational correction terms were usually applied along the principal rotational axes of the main isotopologue. We see no reason why a particular set of axes should be preferred. Problems arise because the second-order correction terms in  $r_m^{(2)}$  fits are only approximately aligned with these principal axes and the Laurie-correction is tied to the X-H bond axes. Corrections are therefore applied along different axes and these axes differ between isotopologues. A choice of coordinate system where many isotopologue principal axes coincide with the  $r_m^{(1)}$  rovibrational correction axes, *i.e.*, the default choice in the literature, is particularly problematic because it amplifies the weight of the remaining asymmetric isotopologues in determining  $r_m^{(2)}$  and Laurie correction parameters. *E.g.*, the unexpected negative Laurie correction term reported for H<sub>2</sub>O by Watson *et al.*<sup>13</sup> is found only when the correction terms are calculated along the main isotopologue principal axes. In the ESI,† section 3.2.5, we describe Python code to systematically explore the angle-dependence of correction terms and fit parameters, using the Monte-Carlo formalism described above.



obtained stable (angle-independent) results with two methods: (a) we randomized the isotopologue coordinate definitions in  $r_m^{(1rL)}$ ,  $r_m^{(2r)}$ , or  $r_m^{(2rL)}$  fits to minimize the correlation between axes for rovibrational and Laurie correction terms. (b) We aligned isotopologue coordinate definitions to be symmetric with respect to the  $a$ - and  $b$ -rotational axes and mapped rovibrational correction terms onto invariant principal axes. Note that method (b) exploits the symmetry of benzene and cannot be generalized to other molecules. In Table 4, rows 8–11, we give values obtained *via* method (b). The full list of fit parameters and parameter correlation tables are given in the ESI,<sup>†</sup> section 3.2.6.

The  $r_m^{(2)}$  parameters should be considered as the best estimate for  $r_e$  equilibrium bond lengths. In this case, two sets of rovibrational correction parameters ( $c, d$  parameters) account for the effective rovibrational displacements of  $r_C$  and  $r_H$ . The  $r_m^{(1)}$  fit only includes first-order rovibrational corrections but gave very similar results. We point out the excellent agreement between the  $r_m^{(2)}$  bond lengths and *ab initio* reference values in Table 1, row 19.

The fitting of an additional Laurie parameter  $\delta_H$  allowed us to explore whether the isotope-dependent bond lengths for C–H and C–D bonds can be distinguished. Unfortunately,  $r_H$  and  $\delta_H$  parameters in  $r_m^{(1L)}$  and  $r_m^{(2L)}$  fits were highly correlated, leading to large parameter uncertainties. This is an obvious sign of an under-determined fit. The  $r_m^{(2L)}$  fit gave a value  $\delta_H = 12 \text{ \AA} \sqrt{u}$ , close to the generally expected value of  $\delta_H \approx 10 \text{ \AA} \sqrt{u}$ . Note that the fit uncertainty for  $\delta_H$  was large and strongly depended on the fitted data-set.

## 4 Discussion

Our high-resolution rotational Raman data for symmetric benzene isotopologues complements high-resolution FTMW data for asymmetrically deuterated isotopologues in the literature. Table 3 gives a summary of benzene isotopologue rotational constants and highlights our contributions. Preceding data for symmetric isotopologues, where available, were based on FTIR or laser excitation spectra with significantly lower resolution.

Mass correlation in our CRASY measurements allowed the unambiguous separation of signals from different molecular species. Up to 5 different spectra (for  $C_6H_6$ ,  $^{13}C-C_5H_6$ ,  $C_6D_6$ ,  $^{13}C-C_5D_6$ , and  $CS_2$ ) were assigned in a single data set. The correlation of spectra in parent and fragment mass channels allowed the direct assignment of fragments to their parent species, confirming expected fragmentation channels.

Fitted isotopologue rotational constants and the corresponding MIM allowed us to derive structure parameters  $r_0$ ,  $r_s$ , and  $r_m$ . A two parameter fit of  $r_0(H)$  ( $=r_0(D)$ ) and  $r_0(C)$  ( $=r_0(^{13}C)$ ) positions reproduced published  $r_0$  estimates with high fidelity. A three parameter fit of distinct  $r_0(H)$ ,  $r_0(D)$ , and  $r_0(C)$  ( $=r_0(^{13}C)$ ), positions indicated a significant isotope effect of  $r_{CH} - r_{CD} > 8 \text{ m\AA}$ . This result is in direct contradiction to a corresponding analysis by Kunishige *et al.*,<sup>15</sup> who found that the H and D bond lengths are identical to within a fraction of a mÅ.

The  $r_0(C)$  bond lengths obtained by two- and three-parameter fits differ considerably (*cf.* Table 4, rows 2 and 5) and only the

former is in agreement with literature estimates.<sup>15,28</sup> The origin of this discrepancy becomes clear when we compare both values to two-parameter fit results based on a data set excluding deuterated isotopologues (*cf.* Table 4, rows 2 and 5 *versus* 3): the assumption of equal bond lengths for C–H and C–D leads to a significant bias in the fit results, which is removed by fitting only heavy carbon isotopologue data or by the three parameter fit. The origin of this bias is obvious: the H,D isotope effect is large due to the large H,D mass difference and the large CH vibrational amplitude. The corresponding  $^{12}C, ^{13}C$  isotope effect is much smaller. The fit results based on  $^{12}C, ^{13}C$  isotopologue data should therefore be considered as best estimate for the  $r_0$  bond lengths.

As outlined in the introduction, it should not be possible to determine isotope-specific bond lengths for all investigated isotope-specific bonds except in particular trivial cases: additional information gained through added isotopic substitution is inherently tied to a corresponding number of additional isotope-dependent geometry parameters. In our three parameter fit, we resolved this issue with the inclusion of  $^{13}C$  substituted isotopologue data and the assumption of  $^{12}C, ^{13}C$  isotopologue-independent bond lengths. By omitting  $^{13}C$  isotopologue data from our fit, we could reproduce the results from Kunishige *et al.*<sup>15</sup> (Table 4, row 6). But in this case the fit parameters became undetermined with uncertainties exceeding the fitted parameter values by factors  $\gg 100$ . The result presented by Kunishige *et al.* therefore represents a highly under-determined solution that is equivalent to the result of the two parameter fit but contains an additional under-determined parameter.

The assumption of  $r_0(^{12}C) = r_0(^{13}C)$  in our three parameter analysis is not exact and explains the overly large apparent H,D isotope effect. Fits with independent  $r_0(^{12}C)$  and  $r_0(^{13}C)$  bond lengths should resolve this issue but run afoul of our caveat that we cannot fit isotope-specific bond lengths for all isotopologues. We nevertheless attempted several four-parameter fits based on further assumptions (details in the ESI<sup>†</sup>) but did not obtain robust geometry parameters. We conclude that the determination of effective isotope-specific  $r_0$  bond lengths based purely on rotational constants is unfeasible, except for the trivial case of linear molecules.

Watson<sup>10,13</sup> introduced explicit assumptions about the mass-dependent scaling of rovibrational corrections, based on the quantum mechanical understanding of rovibrational coupling, which allows us to estimate isotope-independent equilibrium bond lengths. For benzene, results from  $r_m^{(2)}$  fits were very robust and gave bond lengths of  $r_m^{(2)}(C-C) = 1.3913 \text{ \AA}$  and  $r_m^{(2)}(C-H) = 1.0814 \text{ \AA}$ . Fitted bond lengths were stable against exclusion of particular data points,<sup>¶¶</sup> with parameter variations well below 1 mÅ. We therefore consider these values to be the best current estimate for the  $r_e$  bond lengths in benzene.

<sup>¶¶</sup> In particular, we considered excluding all Kunishige values for deuterated benzenes, because their rotational constants were determined with distortion constants  $D_J$ ,  $D_{JK}$ , and  $D_K$  fixed to averaged values from  $C_6H_6$  and  $C_6D_6$ , or excluding our own values for  $^{13}C$  isotopologues, because we determined rotational constants with distortion constants fixed to *ab initio* calculated values.



We must address the significant difference between previous literature estimates for  $r_e$  and our best estimate, based on the  $r_m^{(2)}$  analysis (Table 4, row 8 *versus* 16–18). Pliva *et al.*<sup>28</sup> used a first-order rovibrational correction, based on the determination of a single correction parameter, to estimate equilibrium constants. Kunishige *et al.*<sup>15</sup> applied an identical correction to their results. Our  $r_m^{(2)}$  analysis added second-order rovibrational correction terms, included distinct correction parameters for different rotational axes, and used a significantly larger and more diverse set of rotational data for the analysis. Our  $r_m^{(2)}$  estimate is very close to results from a coupled-cluster reference calculation (Table 4, row 8 *versus* 19).

Additional Laurie-corrections to the  $r_m^{(2)}$  fits greatly increased parameter uncertainties. This is not surprising: three sets of rovibrational correction parameters correct for rovibrational displacements of only two symmetry-unique bonds within the presumed  $D_{6h}$  symmetry. The fitted Laurie parameters and  $r_m^{(2L)}$ (C–H) bond length were no longer robust against exclusion of particular data points. We therefore cannot interpret the fitted Laurie correction terms as a definite estimate for the C–H/D isotope-dependent bond length. But our  $r_m^{(2L)}$  result of  $\delta_H = 12 \text{ \AA} \sqrt{u}$  is close to the generally expected Laurie correction value<sup>57</sup> of  $\delta_H = 10 \text{ \AA} \sqrt{u}$  and corresponds to a reasonable  $\sim 3 \text{ m\AA}$  bond contraction upon H  $\rightarrow$  D substitution.

The propagation of experimental uncertainties with a Monte-Carlo simulation revealed that fitted parameter uncertainties were not due to uncertainties in the experimental data but due to parameter correlation and model errors in our fit models. The rotational data, although measured to high precision, is not sufficient to characterize rovibrational coupling effects and corresponding isotope-specific bond lengths with correspondingly high confidence. This results in parameter uncertainties that are far greater than expected from error propagation.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The authors acknowledge funding support from the National Research Foundation of Korea under Grants No. NRF-2018R1D1A1A02042720 and Samsung Science and Technology Foundation under Grants No. SSTF-BA2001-08.

## Notes and references

- C. Schröter, J. C. Lee and T. Schultz, *Proc. Natl. Acad. Sci.*, 2018, **115**, 5072–5076.
- J. C. Lee, D. E. Lee and T. Schultz, *Phys. Chem. Chem. Phys.*, 2019, **21**, 2857–2860.
- S. Albert, K. K. Albert and M. Quack, in *Handbook of High-resolution Spectroscopy*, ed. M. Quack and F. Merkt, John Wiley & Sons, Ltd, 2011, vol. 2, pp. 965–1019.
- H.-M. Frey, D. Kummli, S. Lobsiger and S. Leutwyler, in *Handbook of High-resolution Spectroscopy*, ed. M. Quack and F. Merkt, John Wiley & Sons, Ltd, 2011, vol. 2, pp. 1237–1265.
- G. Herzberg, F. Patat and J. W. T. Spinks, *Z. Phys.*, 1934, **92**, 87–99.
- J. Kraitichman, *Am. J. Phys.*, 1953, **21**, 17–24.
- C. C. Costain, *J. Chem. Phys.*, 1958, **29**, 864–874.
- D. R. Herschbach and V. W. Laurie, *J. Chem. Phys.*, 1962, **37**, 1668–1686.
- V. W. Laurie and D. R. Herschbach, *J. Chem. Phys.*, 1962, **37**, 1687–1693.
- J. K. Watson, *J. Mol. Spectrosc.*, 1973, **45**, 99–113.
- H. D. Rudolph, *Struct. Chem.*, 1991, **2**, 581–588.
- M. Harmony, V. Laurie, R. Kuczkowski, R. Schwendeman, D. Ramsay, F. Lovas, W. Lafferty and A. Maki, *J. Phys. Chem. Ref. Data*, 1979, **8**, 619–721, DOI: [10.1063/1.555605](https://doi.org/10.1063/1.555605).
- J. K. Watson, A. Roytburg and W. Ulrich, *J. Mol. Spectrosc.*, 1999, **196**, 102–119.
- Z. Kisiel, *J. Mol. Spectrosc.*, 2003, **218**, 58–67.
- S. Kunishige, T. Katori, M. Baba, M. Nakajima and Y. Endo, *J. Chem. Phys.*, 2015, **143**, 244302.
- T. Hirano, U. Nagashima and M. Baba, *J. Mol. Struct.*, 2021, **1243**, 130537.
- B. P. Stoicheff, *J. Chem. Phys.*, 1953, **21**, 1410–1411.
- B. P. Stoicheff, *Can. J. Phys.*, 1954, **32**, 339–346.
- R. C. Lord and T. K. McCubbin, *J. Opt. Soc. Am.*, 1957, **47**, 689–697.
- A. Danti and R. Lord, *Spectrochim. Acta*, 1958, **13**, 180–191.
- R. Anttila, *Z. Naturforsch. A*, 1968, **23**, 1089.
- A. Cabana, J. Bachand and J. Giguère, *Can. J. Phys.*, 1974, **52**, 1949–1955.
- H. B. Jensen and S. Brodersen, *J. Raman Spectrosc.*, 1979, **8**, 103–110.
- J. Kauppinen, P. Jensen and S. Brodersen, *J. Mol. Spectrosc.*, 1980, **83**, 161–174.
- J. Pliva and A. Pine, *J. Mol. Spectrosc.*, 1982, **93**, 209–236.
- J. Pliva and J. W. C. Johns, *Can. J. Phys.*, 1983, **61**, 269–277.
- J. Pliva and A. S. Pine, *J. Mol. Spectrosc.*, 1987, **126**, 82–98.
- J. Pliva, J. Johns and L. Goodman, *J. Mol. Spectrosc.*, 1990, **140**, 214–225.
- J. Lindenmayer, U. Magg and H. Jones, *J. Mol. Spectrosc.*, 1988, **128**, 172–175.
- E. Riedle, T. Knittel, T. Weber and H. J. Neusser, *J. Chem. Phys.*, 1989, **91**, 4555–4563.
- H. Hollenstein, S. Piccirillo, M. Quack and M. Snels, *Mol. Phys.*, 1990, **71**, 759–768.
- M. Junttila, J. Domenech, G. T. Fraser and A. Pine, *J. Mol. Spectrosc.*, 1991, **147**, 513–520.
- J. Domenech, M. Junttila and A. Pine, *J. Mol. Spectrosc.*, 1991, **149**, 391–398.
- M. Okrus, R. Müller and A. Hese, *J. Mol. Spectrosc.*, 1999, **193**, 293–305.
- C. Riehn, A. Weichert and B. Brutschy, *J. Phys. Chem. A*, 2001, **105**, 5618–5621.
- V. Matytilsky, W. Jarzeba, C. Riehn and B. Brutschy, *J. Raman Spectrosc.*, 2002, **33**, 877–883.
- A. Doi, M. Baba, S. Kasahara and H. Katô, *J. Mol. Spectrosc.*, 2004, **227**, 180–186.
- M. Baba, Y. Kowaka, U. Nagashima, T. Ishimoto, H. Goto and N. Nakayama, *J. Chem. Phys.*, 2011, **135**, 054305.



- 39 J. Pliva, A. Valentin, J. Chazelas and L. Henry, *J. Mol. Spectrosc.*, 1989, **134**, 220–226.
- 40 M. Snels, H. Hollenstein, M. Quack, E. Cané, A. Miani and A. Trombetti, *Mol. Phys.*, 2002, **100**, 981–1001.
- 41 J. Pliva, J. Johns and L. Goodman, *J. Mol. Spectrosc.*, 1991, **148**, 427–435.
- 42 M. Oldani and A. Bauder, *Chem. Phys. Lett.*, 1984, **108**, 7–10.
- 43 M. Oldani, R. Widmer, G. Grassi and A. Bauder, *J. Mol. Struct.*, 1988, **190**, 31–40.
- 44 J. Coursey, D. Schwab, J. Tsai and R. Dragoset, in NIST Physical Measurement Laboratory, National Institute of Science and Technology [Online] February 16, 2001, <https://www.nist.gov/pml/atomic-weights-and-isotopic-compositions-relative-atomic-masses>, accessed July 20, 2021.
- 45 W. Gordy and R. Cook, in *Microwave Molecular Spectra*, John Wiley & Sons, New York, 1984, ch. XIII.
- 46 S. Rashev and D. C. Moule, *J. Phys. Chem. A*, 2004, **108**, 1259–1267.
- 47 B. van Eijck, *J. Mol. Spectrosc.*, 1982, **91**, 348–362.
- 48 P. Nösberger, A. Bauder and H. Gunthard, *Chem. Phys.*, 1973, **1**, 418–425.
- 49 R. H. Schwendeman, *Critical evaluation of chemical and physical structural information*, 1974.
- 50 V. Typke, *J. Mol. Spectrosc.*, 1978, **69**, 173–178.
- 51 J. Demaison and H. Rudolph, *J. Mol. Spectrosc.*, 2002, **215**, 78–84.
- 52 C. Schröter, K. Kosma and T. Schultz, *Science*, 2011, **333**, 1011–1015.
- 53 C. M. Western and B. E. Billinghurst, *Phys. Chem. Chem. Phys.*, 2017, **19**, 10222–10226.
- 54 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision A.1*, Gaussian Inc., Wallingford CT, 2009.
- 55 A. Doi, S. Kasahara, H. Katô and M. Baba, *J. Chem. Phys.*, 2004, **120**, 6439–6448.
- 56 *NIST Computational Chemistry Comparison and Benchmark Database*, ed. D. J. Russell, NIST, 2020.
- 57 W. Gordy and R. Cook, in *Microwave Molecular Spectra*, John Wiley & Sons, New York, 1984, ch. XIII, p. 709ff.

