

Twins in rotational spectroscopy: Does a rotational spectrum uniquely identify a molecule?

Marcus Schwarting,¹ Nathan A. Seifert,² Michael J. Davis,³ Ben Blaiszik,⁴ Ian Foster,^{1,4} and Kirill Prozument^{3, a)}

¹⁾*Department of Computer Science, University of Chicago, Chicago, IL 60637, USA*

²⁾*Department of Chemistry and Chemical & Biomedical Engineering, University of New Haven, West Haven, CT 06516, USA*

³⁾*Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439, USA*

⁴⁾*Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439, USA*

(Dated: 26 June 2024)

Rotational spectroscopy is the most accurate method for determining structures of molecules in the gas phase. It is often assumed that a rotational spectrum is a unique “fingerprint” of a molecule. The availability of large molecular databases and the development of artificial intelligence methods for spectroscopy makes the testing of this assumption timely. In this paper, we pose the determination of molecular structures from rotational spectra as an inverse problem. Within this framework, we adopt a funnel-based approach to search for molecular twins, which are two or more molecules, which have similar rotational spectra but distinctly different molecular structures. We demonstrate that there are twins within standard levels of computational accuracy by generating rotational constants for many molecules from several large molecular databases, indicating the inverse problem is ill-posed. However, some twins can be distinguished by increasing the accuracy of the theoretical methods or by performing additional experiments.

Keywords: Inverse problems, rotational spectroscopy, isospectral geometry

I. INTRODUCTION

Pure rotational spectroscopy is a powerful spectroscopic technique in the microwave and millimeter-wave frequency ranges that can reveal detailed structural and dynamical information about a molecule in the gas phase that is not obtainable with other spectroscopic techniques¹. High-resolution rotational spectra distinguish molecular isomers,² conformers,³ and vibrational states,⁴ encode tunneling, fine and hyperfine interactions⁵. The invention of broadband chirped-pulse Fourier transform microwave (CP-FTMW) spectroscopy (also called molecular rotational resonance spectroscopy) enabled fast acquisition of lines over many GHz of spectral bandwidth with meaningful relative intensities while retaining a sub-MHz resolution.^{6,7} The CP-FTMW spectroscopy has opened an opportunity to study chemical reaction dynamics and kinetics by sampling spectra of multiple reaction intermediates with the aforementioned structural and quantum state selectivity.^{8–14} Many of the measured rotational lines remain unassigned, although progress in identifying most of potentially important chemicals exiting a reactor is being made.¹⁵ The need for context-free automated assignment of broadband rotational spectra is even more pressing in analytical chemistry or industrial settings in part because assignment of unknown spectra and identification of molecules that give rise to those spectra require a trained spectroscopist.¹⁶

In the past decade, progress has been made in developing rotational spectroscopy into a practical analytical technique for molecular identification. Many advancements have been

^{a)}Electronic mail: prozument@anl.gov

made in this respect by researchers at BrightSpec company.^{16,17} Though the high spectral resolution of the technique affords the ability to quantify complex mixtures directly in absence of chemical separations,¹⁸ use of gas chromatography¹⁹ can assist in deconvolving complex mixture spectra in a rapid fashion. This methodology has been shown to provide precise enantiopurity measurements of enantiomeric and diastereomeric mixtures without chiral separations, such as those in essential oils¹⁷ and product mixtures from enantioselective hydrogenation reactions²⁰, fluorinated and deuterated active pharmaceutical ingredients,^{21–23} and real-time reaction diagnostics¹⁴. This year, a new acoustic desorption technique has been demonstrated to efficiently volatilize materials up to 600 Daltons,²⁴ which may practically circumvent microwave spectroscopy's largest experimental limitation of requiring sufficient volatilization to enable gas-phase analysis.

Spectral assignment entails attributing experimentally observed spectral lines to transitions between quantum levels with known quantum numbers. That assignment in rotational spectroscopy is based on a quantum mechanical model that adequately describes molecular rotation and intramolecular interactions¹⁶. Determining the parameters in that model, such as the rotational constants, distortion constants, and electric quadrupole coupling constants, is a non-trivial task. In a simple case, a library of reference spectra generated from another experimental study is available. More often, however, lines are not tabulated, and a candidate molecule has to be guessed. Its structure is predicted with *ab initio* methods and the rotational spectrum is simulated. Significant progress has been made toward developing quantum chemistry techniques that provide cost-effective and accurate molecular structures and properties.^{25–27} The resulting $\sim 10^{-3}$ accuracy of predicted transition frequencies, although is an order of magnitude improvement over the standard methods,²⁸ doesn't reach the $\sim 10^{-6}$ experimental accuracy, and fitting of the calculated rotational constants to the measured spectrum is required. Because of that theoretical uncertainty, when the spectral assignment is complete and an experimental spectrum can be reproduced by solving the forward problem, a question may be posed whether another candidate molecule could fit the same spectrum.

The inverse problem in rotational spectroscopy is to identify a molecular geometry either from the set of rotational constants or from the spectrum itself.^{29–31} A solution necessarily exists, but is it unique? In this work we study the latter inverse problem, namely: can a rotational spectrum uniquely define a molecule?

The rest of this paper is as follows. In Section II, we provide background on inverse problems and isospectrality, and on how rotational spectra are analyzed in both the forward and inverse contexts. Next in Section III, we introduce our constructive and exhaustive methods to assess the isospectral nature of rotational spectra. We then present our results in Section IV, first for our constructed environments and then for the datasets we analyzed. Finally, we discuss the implications of these findings when using rotational spectroscopy for sample identification, and consider future directions.

II. BACKGROUND

We first provide a brief introduction to inverse problems and their relevance to spectroscopic analysis, and then review the current state-of-the-art in forward and inverse mapping approaches within rotational spectroscopy.

A. Inverse Problems and Isospectrality in Spectroscopy

Inverse problems can be broadly defined as follows: For some deterministic forward process f (e.g., a dynamical system, machine learning model inference, simulation, or experimental procedure), can one predict f^{-1} , i.e., the input associated with a specific output?³² A natural extension to this question is whether such an inverse mapping from an output to an input is unique. That property of f is known as well-posedness (also called injectivity); a

pair of inputs that result in the same output, and thus demonstrate that f is not well posed, is known as an isospectral collision. Many inverse problems are ill-posed; that is, solutions may be non-unique. In 1966, Mark Kac famously described and explored the inverse problem “can one hear the shape of a drum?”, which poses the question of whether an individual with perfect pitch (capable of accurately describing the entire set of frequencies associated with a sound) can uniquely identify the shape of a drum (defined as a membrane uniformly stretched across a topologically compact region $\Omega \subset \mathbb{R}^2$) by the sounds it produces.³³ For Kac’s query, the forward mapping consisted of applying the Laplacian wave equation across an input surface Ω and identifying nontrivial normal modes through a deterministic process, leading to a discrete series of ordered “tones.” The isospectrality problem of determining whether a given set of tones is unique to an input surface Ω is known to hold for convex surfaces, but counterexamples for concave surfaces have since been identified.³⁴

Outside of acoustics, variations of Kac’s original question have been explored in a variety of domains, including imaging³⁵, signal processing³⁶, photonics³⁷, quantum mechanics³⁸, and spectroscopy³⁹. Many such isospectrality problems in this area of research, commonly known as spectral geometry, remain unsolved or have been solved only under a set of strictly limiting constraints. Furthermore, while Kac and others assume that f may be perfectly observed, in practice whether or not f is well-posed also relies on measurement precision. Two distinct inputs may be distinguishable when measured at higher resolution, but not when measured at lower resolution.

Following the publication of Hückel’s molecular orbital (HMO) theory,⁴⁰ spectral geometry was first employed for chemical systems. Hückel presented a method to compute the molecular orbital $|\psi\rangle$ of π -conjugated systems from a simple linear combination of $2p_z$ atomic orbitals $|\phi_i\rangle$ with corresponding coefficients $\{c_i\}$, written as $|\psi_i\rangle = c_1|\phi_1\rangle + c_2|\phi_2\rangle$. Substituting the above form into the Schrödinger equation, we may write the secular equation $(\mathbf{H} - E\mathbf{S})\vec{c} = 0$ where $\mathbf{S}_{i,j} = \langle\phi_i|\phi_j\rangle$ is the overlap matrix and $H_{i,j} = \langle\phi_i|\hat{H}|\phi_j\rangle$ is the Hamiltonian matrix. Nontrivial eigenvalues from this secular formulation correspond to the respective atomic orbital energies of the system. Günthard and Primas showed how coordinated π -bonds in HMO theory could be represented concisely with a graph adjacency matrix⁴¹ and considered whether distinct molecular graphs representing π -coordinated HMO systems would always have distinct sets of eigenvalues. Collatz and Singowitz first identified isospectral collisions among simple graphs⁴² and many chemically relevant isospectral collisions (or near collisions⁴³) have since been pinpointed^{44–46}. Other molecular representations have also been considered, with Schrier⁴⁷ using a supervised machine learning approach to demonstrate that a set of constitutional isomers of acyclic alkanes cannot be perfectly distinguished by using the Coulomb matrix eigenvalues⁴⁸ as a descriptor.

Inverse and isospectrality problems are increasingly relevant for spectroscopy and analytical chemistry. The analytical power of a spectroscopic technique lies primarily in the degree to which molecules can be uniquely distinguished from one another. When a spectroscopic technique yields results that lead to structural ambiguity, it is common for experimenters to use additional spectroscopic techniques to resolve remaining ambiguity. Even after collecting multiple measurements, some structural ambiguity may remain (such as distinguishing between enantiomers). Nuclear magnetic resonance (NMR) and infrared (IR) spectroscopy are popular analytical techniques because they can rapidly resolve most structural ambiguities to identify a sample. These techniques come with the added bonus that measured spectra can be immediately interpreted to yield structural insights. A significant portion of the inverse problem for small molecules can be performed by human experts or heuristic-based scripts, and interpreting such spectra is a topic covered in most undergraduate chemistry curricula. While rotational spectroscopy greatly surpasses IR spectroscopy in its precision for determining molecular structure in the gas phase, IR spectroscopy is far easier to interpret. Obtaining structural insights from a rotational spectrum alone is not straightforward.

B. Forward and Inverse Mapping in Rotational Spectroscopy

Molecules can be related to their rotational spectra via either a forward mapping (from molecular geometry to spectrum) or an inverse mapping (from spectrum to molecular descriptors). The forward mapping occurs in two steps: the molecule geometry to a set of rotational and dipole constants, and this set of constants to the rotational spectrum.⁴⁹ We term the former the *strong* forward problem and the latter the *weak* forward problem. Likewise, the mapping from rotational spectrum to the set of constants is termed the *weak* inverse problem, and mapping from the set of constants to the molecule geometry is termed the *strong* inverse problem. Figure 1 illustrates the forward and strong/weak inverse problems.

1. Twins and Isospectral Collisions

Uniqueness of the solution for the weak and the strong inverse problems may be expressed with the help of some additional definitions. We assert that two molecules constitute an isospectral collision if they are *indistinguishable* by rotational spectroscopy; that is, if their rotational lines cannot be resolved by a rotational spectrometer. For the microwave region, this normally means that their frequencies are within ~ 10 kHz.

We can also define a looser constraint: we call molecules *twins* if their experimentally measured spectra are distinct from one another, but in the event they are both present in an experimental spectrum, it is unclear which spectrum can be attributed to which molecule.

This distinction arises primarily from the aleatoric uncertainty inherent in aligning simulated molecule constants with experimentally identified molecule constants, a problem that has been studied by Lee and McCarthy²⁸. If a set of indistinguishable molecules can be identified, the inverse problem (weak and strong) is ill-posed. For twin molecules, the weak inverse problem is well-posed (that is, no distinct sets of rotational and dipole constants map to the same spectrum), but the strong inverse problem is generally ill-posed. However, the strong inverse problem can be made well-posed for twin molecules with additional information. This additional information may be obtained by measuring the dipole moment directions derived from relative line intensities, performing higher fidelity simulations of the molecular structure, measuring isotopically-substituted species, collecting Stark or nutation measurements of the dipole moment, or observing intramolecular interactions identified from line splittings. None of these approaches strictly require measurements from a separate spectroscopic technique.

Tackling an inverse problem starts with first defining the forward problem. When considering the forward problem, we assume that an optimized molecule geometry is already available via some classical or quantum mechanical technique. Lee and McCarthy²⁸ present an expected margin of error in DFT-derived rotational and dipole constants compared to experiment, an important consideration when evaluating isospectral constraints. Assuming

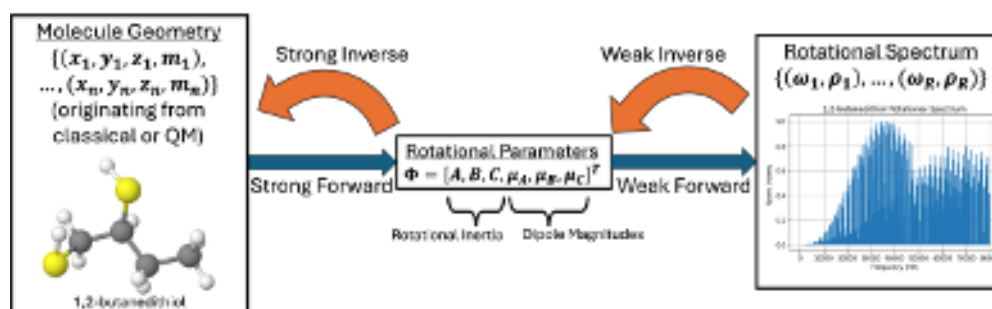


FIG. 1. Diagram showing forward and inverse mapping in rotational spectroscopy.

a pre-computed geometry, we briefly review the strong forward mapping for deriving the rotational and dipole constants of a molecular geometry. (For more information on the formulation of the weak forward mapping, see Gordon, Webb, and Wolpert³⁴ and Kroto⁴⁹.) Finally, we describe current efforts towards efficiently solving both the weak and strong inverse mapping problems.

2. Strong Forward Mapping

A molecule with conformer specificity can be defined by six variables: three rotational constants (A, B, C) and three corresponding dipole constants (μ_A, μ_B, μ_C). Taking the conformer geometry, derived via force field, ab initio, or wave function approaches, as a starting point, suppose a conformer is defined as $\{(m_1, x_1, y_1, z_1), \dots, (m_n, x_n, y_n, z_n)\}$. Then we may define an inertia matrix as

$$\mathbf{A} = \begin{bmatrix} I_{x,x} & I_{x,y} & I_{x,z} \\ I_{y,x} & I_{y,y} & I_{y,z} \\ I_{z,x} & I_{z,y} & I_{z,z} \end{bmatrix}$$

where each element represents a moment of inertia along a pair of Cartesian axes. On-diagonal elements are calculated as

$$I_{x,x} = \sum_{i=0}^n m_i ((y_i - \bar{y})^2 + (z_i - \bar{z})^2)$$

and off-diagonal elements are calculated as

$$I_{x,y} = - \sum_{i=0}^n m_i (x_i - \bar{x})(y_i - \bar{y})$$

which are normalized to the center of mass $(\bar{x}, \bar{y}, \bar{z})$, calculated as

$$\bar{x} = \frac{\sum_{i=0}^n m_i x_i}{\bar{m}}; \quad \bar{m} = \sum_{i=0}^n m_i.$$

Rotational constants (A, B, C) are calculated as $\chi = h/(8\pi^2 I_\chi)$ (where $\chi \in [A, B, C]$, and h is Planck's constant) by using eigenvalues I_χ of the matrix $\mathbf{A}$, ordered as $A \geq B \geq C$. From a dipole vector oriented in Cartesian space (μ_x, μ_y, μ_z) , the rotational dipoles (μ_A, μ_B, μ_C) can be calculated by using a change-of-basis with the rotational eigenmatrix. An important quantity for measuring the degree of asymmetry of a molecule is Ray's asymmetry parameter $\kappa = (2B - A - C)/(A - C)$, where $\kappa = -1$ implies a perfectly prolate rotor and $\kappa = 1$ implies a perfectly oblate rotor. Based on a collection of rotational constants for a diverse set of roughly 400 molecules aggregated by Hellwege and Green⁵⁰, Silbey and Kinsey⁵¹ first observed that most measured molecules were highly prolate. Silbey et al. then derived an equation for the probability distribution of κ based on the construction of a random collection of point masses. As far as we know, this equation is the only attempt to describe the shape distribution of all possible molecular rotors.

3. Inverse Mapping

The weak inverse problem of mapping rotational spectra to a set of rotational and dipole constants remains challenging. However, several semi-automated packages are available to aid researchers. In instances where a subset of rotational transition peaks can be labelled reliably, SPFIT and PGOPHER use a linear least squares procedure to determine the

rotational constants accurately.^{52,53} When a smaller set of transition peaks is available and clear bounds on rotational constants are known, AUTOFIT can determine accurately a set of rotational constants for multiple conformers through a brute-force grid approach,⁵⁴ which has since been scaled to high-performance computing systems⁵⁵. When no transition peaks can be assigned manually, the RAARR package⁵⁶ can mark certain trends in peaks by type (scaffolds) by using heuristics pointed out by Cooke, Ohring *et al.*⁵⁷. However, RAARR requires that strong *a*-type and *b*-type peaks be present in order to construct such trends, and many molecules do not exhibit these peaks. A spectrum of a single molecular carrier may be assigned a set of rotational constants and electric quadrupole constants by using the RAINet artificial neural network.⁵⁸ RAINet is trained on simulated spectra of several classes of molecular rotors (linear, symmetric top, asymmetric *a*-type, *b*-type, *c*-type, with different nuclear spins). Classification and regression take about 200 μ s regardless of the spectral complexity. However, RAINet does not discern spectra from multiple carriers unless it is trained on such mixtures. Other work considers how various distance metrics can be used to measure the space between experimentally observed and computationally proposed spectra.^{59–61}

The strong inverse problem of mapping from rotational and dipole constants to molecular structures is a more daunting challenge. One intuitive approach uses a lookup table to map, for a large set of computed rotational spectra, directly to a molecule identity, thus obviating the need for rotational and dipole constants to be determined.⁶² This lookup approach has successfully been applied to a complex mixture of benzene gasses.¹⁵ McCarthy and Lee³⁰ use a two-step probabilistic deep learning framework to reveal structural information from a set of rotational and dipole constants. First a neural network is employed to predict the largest Coulomb matrix eigenvalues, then a second (probabilistic) neural network is applied to these eigenvalues to predict structural information, including the SMILES string and the presence of various functional groups. However, as Schrier⁴⁷ points out for a set of acyclic alkanes, and as McCarthy and Lee also determine, the lossy Coulomb matrix eigenvalue representation cannot uniquely predict structural information for a molecule. Finally, recent work from Cheng *et al.*³¹ shows how a diffusion-based model can derive structural insights from a set of labelled rotational constants for a parent species and corresponding isotopomeric species by using Kraitchman's equations⁶³ and learning a positive or negative assignment for atomic coordinates.

III. METHODS

In this section we first introduce two constructed environments which generate structures in a constrained and an unconstrained setting. The constrained and unconstrained environments represent two extrema for structural enumeration, with the space of chemically feasible molecular structures situated somewhere between these two extremes. We then describe a funnel process for identifying potential isospectral collisions within large datasets of molecules.

A. Isospectrality by Construction

Suppose we derive rotational constants (A, B, C) for an initial molecule with a given relaxed geometry $\mathcal{M}$. For this fixed structure $\mathcal{M}$, we wish to construct a distinct molecule $\mathcal{M}'$ that is isospectral to $\mathcal{M}$ through an iterative addition of atoms. Note that $\mathcal{M}'$ must not be equivalent to $\mathcal{M}$ via translation, rotation, or reflection, but must possess rotational constants (A', B', C') that are indistinguishable from (A, B, C) . We define *indistinguishable* here to mean that constants are similar to within experimental uncertainty. For small molecules that exhibit only a small number of measurable peaks between 2 and 18 GHz (a common frequency range for structure determination studies that employ microwave spectroscopy⁶⁴), the threshold for experimental uncertainty may be greater than for larger

molecules that exhibit many measurable peaks at fixed intervals based on peak type. We assume that, regardless of the molecules in question, *indistinguishable* implies that frequencies of the pairs of observed spectral lines are within ~ 10 kHz.

When one adds an atom and corresponding bonds to create a new molecule, the geometry must be re-optimized, which affects the Cartesian coordinates of all atoms in the system. In other words, adding a point mass to an existing structure would not guarantee that the re-relaxed structure maintains pairs of observed spectral lines within ~ 10 kHz of the unrelaxed structure. Therefore, such a constructive process could be computationally expensive and may be unlikely to yield the precise collision of rotational constants desired. In contrast to real chemical space, we consider two examples at opposite extremes which do not require relaxation: a constrained environment and an unconstrained environment. The general process of adding atoms to form new valid molecular geometries falls somewhere between these two extremes.

In a constrained environment, structures are defined by a single bond length, a single bond angle, and a single unitary atomic mass. Suppose we begin with a unitary point mass at the origin of $\mathbb{R}^3$ and are permitted iteratively to add unitary point masses only at points that are of unit length away from the existing structure along the x -, y -, or z -axis, with no point masses being placed on top of one another. Such a structure will eventually resemble a square lattice. Supplementary Material Section I elaborates on the construction of such lattices.

Now we consider the opposite extreme: in an unconstrained environment, structures are defined by any arbitrary bond length, any arbitrary bond angle, and any possible atomic mass. Suppose that we begin with a structure $\mathcal{S}$ of n unitary point masses, $n > 2$, with masses $m_1, \dots, m_n \in \mathbb{R}^+$ and each with a different position in Cartesian space:

$$\mathcal{S} = \{(x_1, y_1, z_1), \dots, (x_n, y_n, z_n) | (x_i, y_i, z_i) \in \mathbb{R}^3\}.$$

Suppose further that we can continuously alter the Cartesian coordinates of $\mathcal{S}$ to optimize towards a target $\mathbf{I}_C = (I_{x,x}, I_{x,y}, I_{x,z}, I_{y,y}, I_{y,z}, I_{z,z})$. It is straightforward to identify isospectral collisions in this unconstrained environment. For example, the structure $\mathcal{S}_1 = \{(0, 1, 0), (1, 0, 0), (0, -1, 0)\}$ with corresponding masses $(1, 2, 1)$ is an isospectral collision with $\mathcal{S}_2 = \{(0, 1, 0), (\sqrt{3}, 0, 0), (0, -1, 0)\}$ with corresponding masses $(1, 1, 1)$. We can further justify that, for an arbitrary structure, there are infinitely many distinct structures that are isospectral to the initial spectrum. Finally, we can show that the process of optimizing towards an isospectral collision is nonconvex, with many degrees of freedom allowing for many possible collisions. Supplementary Material Section II elaborates on the construction and optimization of these unconstrained structures.

B. Isospectrality by Molecule Assessment

Rather than attempting to construct geometries from scratch, especially when these geometries may not resemble valid molecules, we also employ large datasets of relaxed molecular geometries to evaluate the potential collisions. Considering relaxed geometries also allows us to evaluate other physical constraints beyond rotational constants that affect the resulting rotational spectrum.

1. The Isospectral Funnel

When evaluating possible collisions, we use a funnel-based approach as shown in Figure 2. Each successive step applies a more rigorous but also more expensive test to remove possible molecule pairs from consideration. With the initial comparisons of rotational constants, we find that $>90\%$ of possible molecule pairs may be removed from consideration. While comparing rotational and dipole constants between a single molecule pair is computationally

inexpensive, the number of considered molecule pairs is large enough to motivate our funnel-based approach to reduce computational overhead.

Starting from a large set of molecule geometries, we begin by enumerating all possible pairs of molecules (M, M') . We then reduce the set of possible pairs based on an overall rotational parameter $R = \sqrt{A^2 + B^2 + C^2}$. We perform a percentage comparison by dividing $|R - R'|$ by $\max(R, R')$, and retain all pairs where the percent difference is $< 1\%$. When molecules are placed in order by an ascending value of R , a percentage comparison is a computationally efficient method to eliminate a significant fraction of possible pairs. Next we compare individual rotational constants, where pairs must satisfy a percent difference of $< 1\%$ for all of (A, B, C) . Another significant fraction of possible pairs can be eliminated in this fashion.

We next perform a similar comparison of dipole ratios, but with several adjustments and edge cases. First, let $r_A = \left(\frac{\mu_A}{\mu'_A}\right)^2$; $r_B = \left(\frac{\mu_B}{\mu'_B}\right)^2$; $r_C = \left(\frac{\mu_C}{\mu'_C}\right)^2$. Next, we normalize these squared dipole ratios according to $\bar{r}_A = \frac{r_A}{\max(r_A, r_B, r_C)}$; $\bar{r}_B = \frac{r_B}{\max(r_A, r_B, r_C)}$; $\bar{r}_C = \frac{r_C}{\max(r_A, r_B, r_C)}$. We then compare pairs of ratios as $\rho_{A,B} = |\bar{r}_A - \bar{r}_B|$; $\rho_{B,C} = |\bar{r}_B - \bar{r}_C|$; $\rho_{C,A} = |\bar{r}_C - \bar{r}_A|$. If $\max(\rho_{A,B}, \rho_{B,C}, \rho_{C,A})$ is less than the specified tolerance, then the pair satisfies our dipole constraint. We specify an absolute tolerance of 0.1 to these dipole ratio comparisons. This metric arises from the assumption that the abundances of the species, n , are not known. Therefore, because the CP-FTMW signal is proportional to $n\mu^{26}$, molecules M and M' cannot be distinguished by the ratios $\left(\frac{\mu_x}{\mu'_x}\right)^2$. However, the difference in relative intensities of the a -type and b -type transitions within each spectrum, for example, are observable and expressed here through $\rho_{A,B}$.

Since experimental measurements of rotational spectra are extremely accurate in the frequency domain (with peak frequencies arising solely from the contribution of rotational constants) and are less accurate in the intensity domain (with peak intensities arising from the contribution of both rotational and dipole constants, and further complicated by non-equilibrium experimental conditions,⁶⁵ and imperfections in the apparatus calibration), we assign the tolerances of rotational and dipole constants according to this difference in simulation and detection accuracy. There are several edge cases with respect to this approach of comparing dipole magnitudes. First, we presume that any dipole magnitude $< 0.05D$

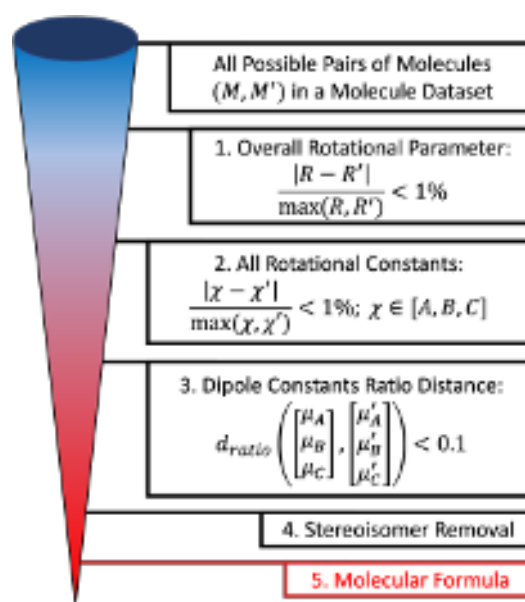


FIG. 2. Funnel diagram for evaluating possible spectral twins in a dataset of molecule geometries.

cannot be readily measured experimentally. We also presume that any dipole magnitude $> 0.1D$ can be measured experimentally. For a pair of molecules, suppose $\mu_\chi < 0.05$ and $\mu'_\chi > 0.1$. Then this pair cannot be a set of twins, because the first species exhibits no χ -type peaks, while the second species exhibits χ -type peaks. If instead $\mu_\chi < 0.05$ and $0.05 < \mu'_\chi < 0.1$, then we employ the standard ratio approach to compare these dipoles. If rather $\mu_\chi < 0.05$ and $\mu'_\chi < 0.05$ (which would occur for a planar or near-planar molecule), then we omit the r values that include this dipole component, leaving the remaining value of ρ which does not include the χ dipole component as the only source of comparison for the tolerance. Finally, suppose $\mu_{\chi_1} < 0.05$; $\mu_{\chi_2} < 0.05$ and $\mu'_{\chi_1} < 0.05$; $\mu'_{\chi_2} < 0.05$ (which would occur for a linear or near-linear molecule). Then, so long as the remaining dipole component is measurable for both species, the pair of molecules are within tolerance by default.

Next, we remove molecule pairs which are stereoisomers of one another. We use the MolVS package⁶⁶ to remove stereoisomer pairs, and further convert structures to canonical SMILES strings which are agnostic to chirality⁶⁷. Finally, we provide an optional comparison based on molecular formula. While a molecular formula cannot be immediately determined from a rotational spectrum alone (as this is part of the strong inverse problem), it can be easily assessed by using mass spectrometry,⁶⁸ which can be run in conjunction with rotational spectroscopy.

The funnel we have so far defined considers only a single conformer of a molecule without isotopic substitution. In the case where multiple relaxed geometry conformers (and corresponding relative energies) are available per molecule, the same funnel process described above can occur by using the lowest-energy conformer of each pair, but with a final step that requires a less rigorous alignment for pairs of available higher energy conformer geometries. Similarly, an additional step in the funnel might consider spectral differences caused by isotopic substitution. Rejecting pairs based on isotopic substitution assumes that the experimental signal-to-noise ratio (SNR) exceeds 300:1 in the case of ^{13}C substitution, 1500:1 in the case of ^{18}O substitution, and roughly 18000:1 in the case of deuterium substitution (assuming natural abundance without enrichment). While a majority of instruments can capture ^{13}C substitution without difficulty, most instruments cannot easily capture deuterium substitution at natural abundance. Since experimental SNR can vary widely by instrument, pre-sets, and measurement time, we do not include differences arising from the presence of isotopomers in rotational spectra as a source of further disambiguation, although we acknowledge that it could be used as such. The same may be said for other measurable effects such as hyperfine interaction and internal rotation.

2. Molecule Datasets

We consider several molecule datasets spanning a range of molecule size, diversity, and geometry fidelity: see Table I. **QM9**⁶⁹ comprises 1.33×10^5 chemically valid structures with up to nine (C,O,N,F) heavy atoms (and implicit H atoms), providing for each a DFT-optimized geometry (using B3LYP/6-31G(2df,p)) and corresponding Cartesian-oriented dipole vector. **QM7x**⁷⁰ includes 6950 chemically valid structures containing up to seven (C,N,O,S,Cl) heavy atoms (with implicit H atoms). Unlike QM9, QM7x provides multiple DFT-optimized geometries and a Cartesian-oriented dipole vector for each molecule (using PBE0+MBD), for a total of 4.03×10^6 DFT-optimized conformers. Thus each unique molecule in QM7x has an average of 580 distinct conformers.

Enumerating the conformational diversity of a set of large molecules by using high-fidelity DFT is computationally expensive. Thus, larger datasets of chemically diverse molecules are commonly enumerated by using lower-fidelity approaches. The GEOM dataset⁷¹ employs XTb-CREST⁷² to cheaply enumerate and relax multiple low-energy conformers of large molecules. This dataset can be split into two parts. **GEOM-QM9** contains the same set of molecules as QM9, but enumerates a total of 1.82×10^6 conformers by using XTb-CREST, for an average of roughly 14 conformers per molecule. **GEOM-Drug** comprises 2.91×10^5

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0212632

TABLE I. Properties of the seven molecule datasets considered in this work.

Dataset	Molecules	Conformers	Theory	Average Mol. Wt.	Avail. SMILES	Avail. Dipoles	Ref.
QM9	1.33×10^5	1.33×10^5	B3LYP	122.69 Da	Yes	Yes	⁶⁹
QM7x	6950	4.03×10^6	PBE0+MBD	96.58 Da	No	Yes	⁷⁰
GEOM-QM9	1.33×10^5	1.82×10^6	XTB-GFN2	122.69 Da	Yes	Yes	⁷¹
GEOM-Drug	2.91×10^5	3.12×10^7	XTB-GFN2	355.24 Da	Yes	Yes	⁷¹
GEOM-Drug (Top 1)	2.91×10^5	2.91×10^5	XTB-GFN2	355.24 Da	Yes	Yes	⁷¹
PubChem	$>1.10 \times 10^8$	$>1.10 \times 10^8$	Varied	420.97 Da	Yes	No	⁷³
PubChem ($W_{LH} \pm 10$)	6.78×10^6	6.78×10^6	Varied	254.07 Da	Yes	No	⁷³
NearIB	8621	8621	Varied	237.65 Da	Yes	Yes	⁷³

drug-like molecules identified across several sources, and enumerates a total of 3.12×10^7 conformers, for an average of 107 conformers per molecule. For our isospectral evaluation, we consider only the lowest-energy conformer available per molecule in GEOM-Drug. The GEOM dataset does not include dipole calculations, therefore in instances where a collision is deemed possible (based on R, A, B, C), we perform an XTB-GFN2 point calculation by using the available GEOM coordinates to determine and compare these principal axis oriented dipoles.

We draw all available geometries from **PubChem**, which totals over 110 million unique molecules, with one geometry per molecule and dipole moments generally unavailable⁷³. To the best of our knowledge, the PubChem dataset represents the largest set of aggregated molecule geometries currently available. The fidelities of these geometries may vary widely, ranging from high-fidelity DFT approaches to simple force field relaxations. We select from PubChem all molecules within a certain molecular weight range, which we obtain based on the hypothesized high structural diversity of molecules in the range, as we now describe. Lüttschwager *et al.*⁷⁴ perform XTB simulations on successively longer alkane chains and find that at $C_{18}H_{38}$ (with a weight of $W_{LH} = 254$ Da), the alkane chain does not uphold a *trans*- orientation across all carbon-carbon bonds as the lowest-energy state, but instead takes a *cis*- orientation on a middle carbon-carbon bond, forming a hairpin as the new lowest-energy state. The transition of molecules with repeating subunits from a highly prolate configuration (*trans*- oriented bonds only) to a more spherical configuration (with some *cis*- oriented bonds) would suggest a high degree of structural diversity. We therefore select a subset of PubChem structures in the range $W_{LH} \pm 10$ Da [i.e., (244, 264) Da], and only assess potential collisions on R, A, B , and C . Finally, we select a small subset of molecules from structures that includes the drug Ibuprofen along with 8620 molecules the closely resemble Ibuprofen (in terms of Tanimoto similarity). We then calculate the dipole components of this subset with an XTB point calculation. We refer to this subset as the NearIB dataset.

IV. RESULTS

We first consider the results of our constrained and unconstrained constructive geometries, then review the outcome of our isospectral funnel applied across molecular datasets.

A. Constructive Isospectrality Approaches

First, we consider the inherent difficulty with generating isospectral pairs in the constrained environment. This difficulty can be attributed to the combinatorial explosion of possible structures for the given set of point masses. Our analysis, further described in Supplementary Material Section I, never uncovered an isospectral pair of any size (that

was not isomorphic with respect to a translation, rotation, or reflection) in the constrained environment, and it is unclear whether such an isospectral pair could be constructed (either from structures in $\mathbb{R}^3$, or in $\mathbb{R}^n$; $n \geq 2$).

Compared to the constrained environment, generating twins to arbitrary numerical precision is straightforward in the unconstrained environment. We also find that structures are not required to have the same number of point masses to identify isospectral collisions in an unconstrained environment, so long as the number of point masses exceeds three. Furthermore, an arbitrary number of distinct isospectral collisions can be achieved through various numerical approaches described in Supplementary Material Section II.

B. Molecule Analysis

We begin by considering the distribution of molecules across datasets by using Ray's κ , then by exploring the isospectral collisions identified for each dataset.

1. Dataset Summary

Figure 3 shows the cumulative distribution of Ray's κ across both constructed environments and molecule datasets, and also the theoretical cumulative distribution of κ derived by Silbey and Kinsey⁵¹ from a closed-form expression based on an assumption of uniformly distributed moments of inertia among molecules. The κ distribution across all datasets shows that, in general, small to mid-sized organic molecules lean heavily towards a prolate rotor. This prolate proclivity confirms earlier observations from Silbey and Kinsey⁵¹, however the tendency for both random constrained structures and real molecules (QM9 and GEOM-Drug) to remain highly prolate is even more extreme than they first suggested. Interestingly, we see that the distribution across κ for real molecules more closely resembles the κ distribution of the constrained environment than the unconstrained environment.

Figure 4 shows a box-and-whisker plot of Ray's κ across the set of PubChem molecules, separated by ranges in molecular weight. Silbey's distribution is also shown with labelled quartiles. The distribution of PubChem molecular weights within these ranges is available in Supplementary Material Section III. The distribution of structures between 0 and 100 Da is almost entirely prolate, but structures become even more concentrated at a prolate extreme at higher molecular weights. Note that outliers are not included in Figure 4. We see an inflection in the 300 to 400 Da range (with average $\kappa \approx -0.92$), which is the most prolate range, after which molecules become more asymmetric (and even oblate). The set of structures with molecular weight >1000 Da contains many oblate structures, and has an average $\kappa \approx -0.32$.

In Supplementary Material Section V, we briefly compare the distribution of rotational constants and κ between the lowest-energy conformers GEOM-QM9 dataset⁷¹ and those in the high-fidelity DFT-optimized QM9 dataset⁶⁹. We see a shift towards a more prolate extreme with high-fidelity DFT geometries, although it is unclear whether this trend continues at even higher-fidelity DFT methods.

2. Molecule Isospectral Analysis

Table II shows the number of possible twin pairs (within pre-defined tolerances) across molecule datasets along each step of the funnelling process. For datasets containing only a single conformer per distinct molecule (QM9, GEOM-Drug, and PubChem), we find that funnelling on R results in about 1/100 remaining pairs. For QM9 and GEOM-Drug, we see that collisions on rotational constants are rare (roughly 1/20000). When many conformers of the same molecule are considered (as with QM7x and GEOM-QM9), many more matches on rotational constants occur. These collisions among structures can come

TABLE II. Remaining twins from isospectral funnelling of conformers across various datasets.

Dataset	Considered Conf. Pairs	Overall Rot. Collision	Rot. Const. Collision	Dipole Collision	Stereo. Collision	Formula Collision
QM9	8.45×10^9	2.34×10^8	3.35×10^5	356	295	36
QM7x	8.11×10^{12}	1.91×10^{11}	3.46×10^8	1.38×10^6	—	1.89×10^5
GEOM-QM9 (All)	1.66×10^{12}	4.35×10^8	5.85×10^5	5.98×10^3	349	27
GEOM-Drug (Top 1)	4.23×10^{10}	4.90×10^8	4.25×10^5	941	621	462
PubChem ($W_{LH} \pm 10$)	2.30×10^{13}	3.23×10^{11}	6.36×10^9	—	—	—
NearIB	3.72×10^7	4.97×10^5	2.03×10^3	766	404	404

from conformers of either the same or different molecules. Comparing collisions for dipole moment ratios narrows the set of possible pairs once more: for datasets with a single conformer per molecule, the pair reduction is roughly a further three orders of magnitude. Figure 5 shows the effect of changing the tolerance on R and (A, B, C) over the range 0.01% to 1%. If rotational and dipole constants could be obtained via simulation at an even higher level of accuracy, the number of remaining twin pairs could decrease yet further.

Figure 5 shows the density of twins based on rotational constants with respect to the molecular weight. Here the weight of a twin pair is an average of the two molecular weights. We find that the fraction of twins, normalized by the density of all molecules in QM9, peaks at around 120 Daltons. The decay of twin density beyond 120 Daltons is unexpected, however this may be related to the declining density of overall molecule weights in the corresponding QM9 database. We also find that the difference in molecular weights within a pair of twins can be separated by as much as 30 Daltons.

We now consider a number of twins across each dataset to characterize the sorts of molecules which remain after funnelling. Figure 6 shows an example of twins (on rotational

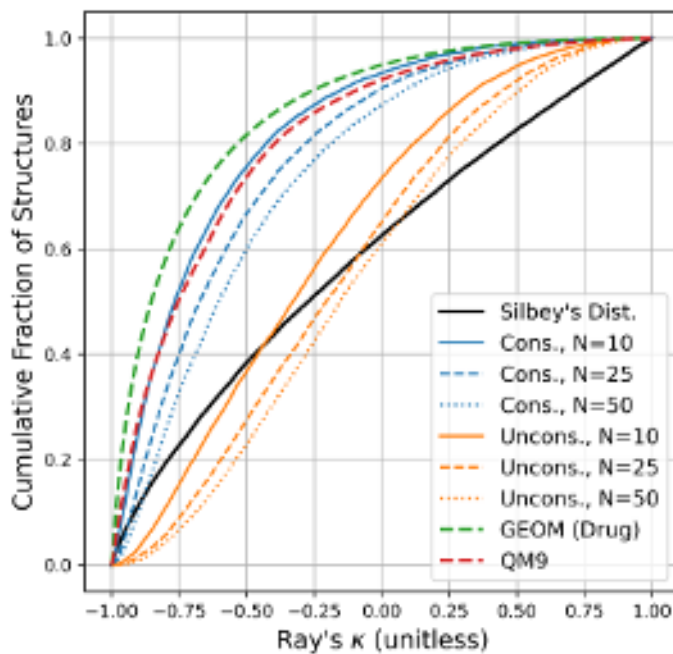


FIG. 3. Cumulative distribution of Ray's κ across constrained (blue) and unconstrained (orange) environments with a varying number of point masses (10, 25, 50). Silbey's cumulative probability distribution on κ is also shown (black), along with the cumulative distribution of κ within GEOM (Drug) and QM9 datasets (green and red, respectively).

constants only) that are separated by a large difference in molecular weight, demonstrating that two molecules do not always have to be of a comparable mass to be considered twins. Figure 7 shows two examples of structural isomers from QM9 with very similar rotational constants and a single strong dipole component (μ_B). Note that ^2H or ^{15}N isotopic substitution on each species would produce distinct rotameters which, once labelled, would remove structural ambiguity. Figure 8 shows examples of twins which are not structural isomers in QM9, and could in practice be distinguished by using mass spectrometry. If such a measurement cannot be taken, these molecules are also more conformationally flexible and could therefore be distinguished by the presence of other, distinct low-energy conformers. For a case where both species could be present in the same sample, a nutation experiment could provide distinguishing information about the relative dipole intensities of both species (without performing a Stark measurement). Figure 9 shows examples of twin conformers in QM7x, which occur alongside a number of other conformers for both species. In practice these spectra are distinct when multiple conformers are present, with the exception of this pair of twin conformers.

Finally, Figure 10 shows examples of molecules from the GEOM-Drug dataset. It is unclear whether these molecules could be observed via rotational spectroscopy, even when using ablation techniques. Furthermore, we did not identify any twins associated with Ibuprofen, although twins do exist between pairs of molecules other than Ibuprofen in the NearIB dataset. Figure 11 shows three molecules with similar rotational and dipole constants which constitute the nearest matches to Ibuprofen.

V. DISCUSSION

We assert that, with respect to structural diversity, the space of possible molecules with their conformers lies somewhere between the two constructed extremes we have presented. Of course, atoms in a molecule are not constrained to a single discrete mass, bonded atom pairs may be of varying lengths, and bond angles are not limited to 90° . Conversely, atoms in a conformer are not fully unconstrained with regard to masses, bond lengths, and

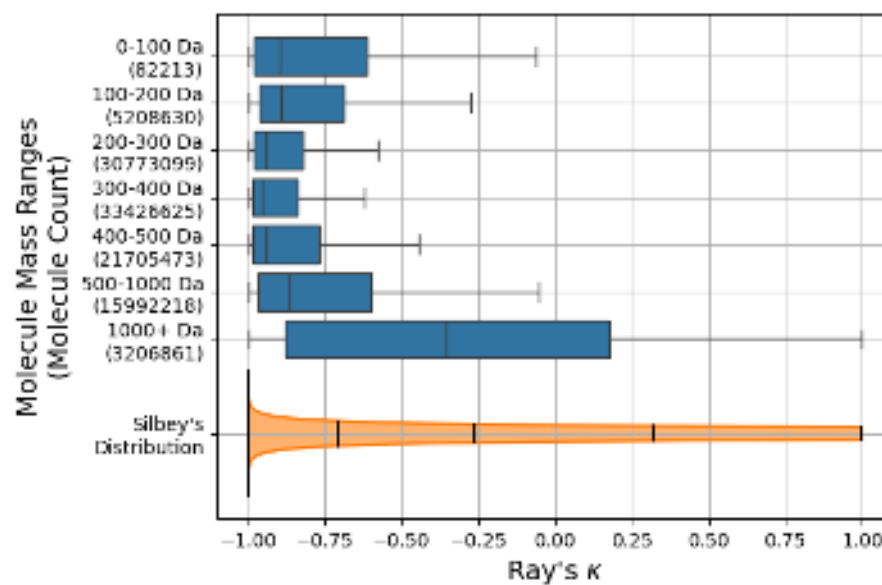


FIG. 4. Box-and-whisker distribution of Ray's κ for all molecules in the PubChem dataset, for binned masses in Daltons (Da). The counts of PubChem molecules falling into each mass range is also listed. Silbey and Kinsey's probability distribution on κ is also shown⁵¹, with black lines representing corresponding quartiles.

bond angles. It is unclear whether the properties of the space of conformers more closely resemble the constructed constrained or the unconstrained environment. However, based on the distribution of Ray's κ in Figure 3, the space of molecule conformers seems to resemble the constrained environment in terms of κ .

The forward mapping of structures in the constrained environment to moments of inertia appears well-posed, since no isospectral collisions were observed. By contrast, the function mapping structures in the unconstrained environment to moments of inertia is ill-posed,

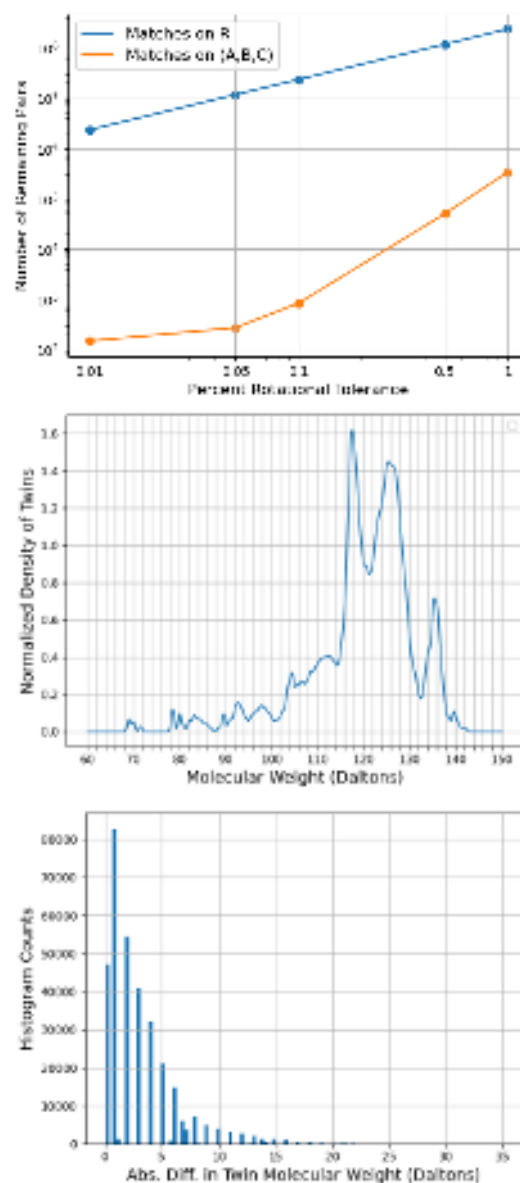


FIG. 5. Top: The remaining number of twin pairs in QM9 based on the overall rotational inertia R as well as rotational constants (A, B, C). The number of remaining twin pairs becomes progressively smaller as tolerance decreases from 1% to 0.01%. Middle: Density of twins (based on rotational constants only) in QM9 with respect to molecular weight, normalized by the overall molecule density in QM9 with respect to molecular weight. Bottom: Histogram of twins (based on rotational constants only) in QM9 with respect to the difference in molecular weights between twins.

with an infinite number of distinct structures capable of satisfying the same set of moments of inertia. If the space of molecules resembles the constrained environment, as suggested by Figure 3, this may imply that no indistinguishable collision exists.

From our analysis of conformer-specific molecular datasets, we find that many pairs of molecules have rotational constants within our specified tolerances. However, far fewer twin pairs are identified that have both rotational and dipole constants within our specified tolerances. The examples in Figure 7 are among the closest matches between molecules in QM9, and both pairs of similar structures offer little conformational flexibility. Supposing these molecules were present in a mixed sample, experimentally their spectra could likely be distinguished from one another quite readily, although their respective identities may remain ambiguous. As previously mentioned, it is common for rotational spectrometers to assess peak frequencies with ~ 10 kHz accuracy, which corresponds to the relative uncertainty in measured rotational constants of $\Delta\chi/\chi \sim 10^{-4} - 10^{-6}$, $\chi \in [A, B, C]$. At the same time, the DFT (B3LYP/6-31G)-calculated uncertainty is $\Delta\chi/\chi \sim 10^{-2}$.²⁸ In other words, two molecules may be twins if they are experimentally distinguishable, but the discrepancy between measurement and simulation prevents us from accurately identifying which species aligns with which peaks in an experimental spectrum.

The examples from QM7x in Figure 9 show another pair of twin molecules. However, in this case the comparison is between two conformers of different molecules where each molecule exhibits many possible conformers. In practice, these molecules could be distinguished from one another by the presence of other conformers. In all cases presented, isotopomeric species can play a part in distinguishing these molecules as well. Thus, even in instances where near-isospectral matches between molecules or conformers appear in the structure datasets we evaluated, the pairs may be distinguished by using some combination of nutation, conformational flexibility, isotopomeric substitution, or molecular formula information from mass spectrometry. Instances where this additional level of analysis is required to distinguish between species appear to be rare in practice.

We also consider the number of remaining twins for molecules of varying masses. We compare QM9, with average molecular weight of 122.69 Da, and GEOM-Drug (Top 1), with average molecular weight of 355.24 Da, to see how changes in mass influences the number of remaining twins. Even though the set of conformer pairs in GEOM-Drug (Top 1) was

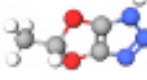
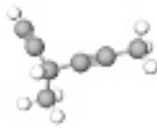
Category	Molecule 1	Molecule 2
Lewis Structure		
R, MHz	4948.75	4970.63
A, MHz	4611.44	4631.89
B, MHz	1393.04	1403.50
C, MHz	1133.22	1132.71
μ_A^2, D^2 (ratio)	3.01 (1.00)	0.21 (0.95)
μ_B^2, D^2 (ratio)	0.19 (0.06)	0.15 (0.68)
μ_C^2, D^2 (ratio)	0.78 (0.26)	0.22 (1.00)
Mol. Formula	$C_4H_5N_3O_2$	C_7H_8
Mol. Weight (Da)	92.14	127.10

FIG. 6. A pairs of molecules from the QM9 dataset that are twins (by rotational constants only). These molecules have the greatest separation in molecular weights out of all twin molecules in QM9 (by rotational constants only). The rotational constants, squared dipole constants, squared dipole ratios (relative to the largest dipole), and molecular formulae are listed for each. Their SMILES are CC1Oc2nn[nH]c2O1 and C#CC(C)C#CC, respectively.

roughly five times the size of the set of conformer pairs in QM9, the number of remaining twin pairs after funnelling by R and (A, B, C) is a smaller fraction of total possible conformer pairs. Since rotational spectroscopy is a gas-phase technique, compounds with low volatility or high boiling points may not be measurable. While this may imply a ceiling on the size of molecules that can be analyzed by using rotational spectroscopy, a number of ablation techniques have been devised to coax large aromatic molecules into the gas phase.^{75,76} Recently, a promising acoustic desorption technique has been demonstrated to put larger species in the gas phase.²⁴ Molecular complexation can also be measured for a number of species by using rotational spectroscopy at specific conditions, with correspondingly low rotational constants permitting many molecule complex conformers to be uniquely identified.⁷⁷

VI. CONCLUSION AND FUTURE WORK

Advancement of broadband rotational spectroscopy to the realms of chemical kinetics and analytical chemistry relies on its capacity to discern robustly between distinct molecules. This discerning power can be mathematically expressed through the well-posedness of the inverse problem that maps spectra to molecular structures. While we know that a molecular structure that gives rise to a spectrum exists, in this work we have explored whether the structure that produces such a spectrum is unique, which would make the inverse problem well-posed.

First, we construct constrained and unconstrained environments and assess how isospectral collisions—the instances of different molecular structures having an indistinguishable set of rotational constants—can be identified. We find that (contrived) *constrained* environments produce structures more similar to real molecules (according to Ray's κ), and do not yield any isospectral collisions. In contrast, the spatially *unconstrained* assembly of point masses readily leads to collisions with arbitrary numerical precision.

Second, we search several large datasets of calculated molecular geometries for potential isospectral collisions by using a funnelling approach. The number of collisions falls rapidly as the number of parameters to be matched (such as rotational constants and dipole moment projections) increases, and as the allowed uncertainties in these parameters are tightened.

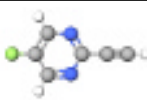
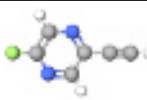
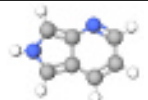
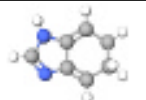
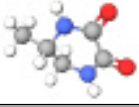
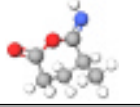
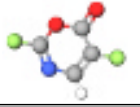
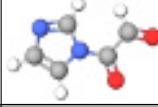
Category	Molecule 1	Molecule 2	Category	Molecule 1	Molecule 2
Lewis Structure			Lewis Structure		
R, MHz	6202.25	6203.50	R, MHz	4440.80	4444.52
A, MHz	6055.05	6056.74	A, MHz	3943.82	3942.30
B, MHz	1020.53	1019.03	B, MHz	1670.38	1680.43
C, MHz	873.34	872.27	C, MHz	1173.39	1178.21
μ_A^2, D^2 (ratio)	0.71 (1.00)	1.35 (1.00)	μ_A^2, D^2 (ratio)	11.52 (1.00)	11.41 (1.00)
μ_B^2, D^2 (ratio)	<0.01 (<0.01)	<0.01 (<0.01)	μ_B^2, D^2 (ratio)	0.01 (<0.01)	0.01 (<0.01)
μ_C^2, D^2 (ratio)	<0.01 (<0.01)	<0.01 (<0.01)	μ_C^2, D^2 (ratio)	<0.01 (<0.01)	<0.01 (<0.01)
Mol. Formula	$C_6H_3N_2F$	$C_6H_3N_2F$	Mol. Formula	$C_7H_6N_2$	$C_7H_7N_2$

FIG. 7. Two pairs of molecules from the QM9 dataset that are twins. The rotational constants, squared dipole constants, squared dipole ratios (relative to the largest dipole), and molecular formulae are listed for each. For the pair of molecules on the left, their isomeric SMILES are C#CC1=NC=C(C=N1)F and C#CC1=CN=C(C=N1)F, respectively. For the pair of molecules on the right, their isomeric SMILES are C1=CC2=CNC=C2N=C1 and C1CC=C2C(=C1)NC=N2, respectively.

We found instances of molecule twins, which have predicted rotational and dipole constants close enough that a standard molecular simulation would not be able to discern which spectrum corresponds to which of the two molecules, even if their spectra were measured to be distinct. It is possible that with higher accuracy calculations of zero-point-averaged molecular structures or with additional experiments (such as isotopic substitution or nutation) these collisions could be resolved.

We conclude that for molecules in the present datasets, the mapping from spectra to structures may be well-posed in principle, but is ill-posed at the current level of accuracy offered by reasonably fast calculations of structures. Although we only identify twin pairs in this work, it remains unclear whether any pairs of distinct molecules (that are not enantiomers) with experimentally indistinguishable rotational spectra will be identified.

Several other possibilities could be considered when assessing isospectrality constraints. First, Boltzmann-weighted conformer abundances may help identify chemical species for which spectra of multiple conformers in a thermodynamic equilibrium are measured. Next, rather than comparing only the constants pertaining to specific structures, another framing

Category	Molecule 1	Molecule 2	Category	Molecule 1	Molecule 2
Lewis Structure			Lewis Structure		
R, MHz	3345.83	3347.47	R, MHz	3706.90	3709.82
A, MHz	2889.16	2890.67	A, MHz	3286.84	3288.60
B, MHz	1381.84	1379.27	B, MHz	1403.67	1406.25
C, MHz	968.43	973.24	C, MHz	983.61	985.04
μ_A^2, D^2 (ratio)	19.78 (0.51)	4.36 (0.67)	μ_A^2, D^2 (ratio)	3.46 (1.00)	1.92 (1.00)
μ_B^2, D^2 (ratio)	<0.01 (<0.01)	<0.01 (<0.01)	μ_B^2, D^2 (ratio)	0.75 (0.22)	0.44 (0.23)
μ_C^2, D^2 (ratio)	31.07 (1.00)	6.48 (1.00)	μ_C^2, D^2 (ratio)	<0.01 (<0.01)	<0.01 (<0.01)
Mol. Formula	$C_5H_8N_2O_2$	$C_6H_8NO_2$	Mol. Formula	$C_4HNO_2F_2$	$C_5H_4N_2O_2$

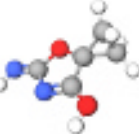
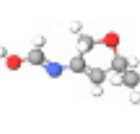
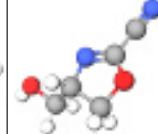
Category	Molecule 1	Molecule 2	Category	Molecule 1	Molecule 2
Lewis Structure			Lewis Structure		
R, MHz	3495.14	3500.88	R, MHz	3647.35	3656.35
A, MHz	2911.05	2916.15	A, MHz	3396.78	3405.40
B, MHz	1580.45	1586.37	B, MHz	1035.75	1037.58
C, MHz	1115.33	1111.62	C, MHz	832.05	834.03
μ_A^2, D^2 (ratio)	0.65 (1.00)	0.01 (1.0)	μ_A^2, D^2 (ratio)	0.15 (0.04)	0.28 (0.03)
μ_B^2, D^2 (ratio)	0.22 (0.34)	<0.01 (0.32)	μ_B^2, D^2 (ratio)	3.52 (0.85)	6.51 (0.82)
μ_C^2, D^2 (ratio)	<0.01 (<0.01)	<0.01 (0.07)	μ_C^2, D^2 (ratio)	4.16 (1.00)	7.98 (1.00)
Mol. Formula	$C_5H_6N_2O_2$	C_8H_{16}	Mol. Formula	$C_6H_9NO_2$	$C_5H_6N_2O_2$

FIG. 8. Four pairs of molecules from the QM9 dataset that are twins. The rotational constants, squared dipole constants, squared dipole ratios (relative to the largest dipole), and molecular formulae are listed for each. For the pair of molecules on the top left, their canonical SMILES are CC1CNC(=O)C(=O)N1 and CC1CCC(=O)OC1=N, respectively. For the pair of molecules on the top right, their canonical SMILES are O=c1oc(F)ncc1F and O=CC(=O)n1ccn1, respectively. For the pair of molecules on the bottom left, their canonical SMILES are N=C1N=C(O)C2(CC2)O1 and CC1CCCC(C)C1, respectively. For the pair of molecules on the bottom right, their canonical SMILES are CC1C=C(N=CO)CO1 and N#CC1=NC(CO)CO1, respectively.

of the isospectrality question might compare generated spectra by using an optimal transport distance, Hamming, or Minkowski $P \neq 2$ metric. Further, additional constants can be considered which more subtly affect a measured rotational spectrum, including quartic centrifugal distortion constants and nuclear electric quadrupole coupling constants. These constants could be considered alongside rotational and dipole constants in the future. Finally, assessing larger datasets of geometries and conformers might still uncover twin conformers that are irreconcilable from either a theoretical or an experimental perspective. Twins may also be identifiable among distinct complexes of molecules, which may exhibit rotational symmetries that single molecules cannot emulate. In this work we limit our consideration to pure rotational spectroscopy, but the similar questions are equally valid for other spectroscopic modes.

SUPPLEMENTARY MATERIAL

Supplementary Material contains details on methods and results related to the search for isospectral collisions in the constrained and unconstrained environments. Additional analysis of the distributions of molecular properties in the PubChem and QM9 datasets is also available.

DATA AVAILABILITY

The code associated with the findings in this paper are available via GitHub: [\[1\]](#).

AUTHOR DECLARATIONS

K. P. holds patents US11380422B2 and US11594304B2 that are related to this work.

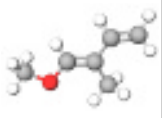
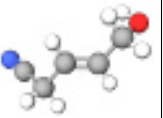
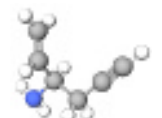
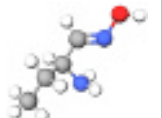
Category	Molecule 1	Molecule 2	Category	Molecule 1	Molecule 2
Lewis Structure			Lewis Structure		
R, MHz	7294.63	7294.79	R, MHz	4398.18	4398.72
A, MHz	7095.97	7096.18	A, MHz	3654.51	3654.97
B, MHz	1281.40	1281.15	B, MHz	1890.04	1890.24
C, MHz	1103.10	1103.13	C, MHz	1554.46	1554.63
μ_A^2, D^2 (ratio)	0.13 (1.00)	0.58 (1.00)	μ_A^2, D^2 (ratio)	0.02 (1.00)	0.03 (1.00)
μ_B^2, D^2 (ratio)	0.03 (0.19)	0.10 (0.18)	μ_B^2, D^2 (ratio)	<0.01 (0.05)	<0.01 (0.05)
μ_C^2, D^2 (ratio)	<0.01 (0.02)	<0.01 (<0.01)	μ_C^2, D^2 (ratio)	<0.01 (0.03)	<0.01 (0.02)
Mol. Formula	$C_6H_{10}O$	C_5H_7NO	Mol. Formula	C_6H_9N	$C_4H_{10}N_2O$

FIG. 9. Two pairs of molecule conformers from the QM7x dataset that are twins. The rotational constants, squared dipole constants, squared dipole ratios (relative to the largest dipole), and molecular formulae are listed for each. For the pair of molecules on the left, their isomeric SMILES are C/C(=C\OC)/C=C and C(/C=C/CO)C#N, respectively. For the pair of molecules on the right, their isomeric SMILES are C=CC(CC#C)N and CCC(/C=N\O)N, respectively.

ACKNOWLEDGMENTS

The authors thank Logan Ward for his thoughts, comments, and ideas on isospectrality in spectroscopy. This material is based on work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences, and by the DOE Office of Advanced Computing Research, under Contract DE-AC02-06CH11357.

- ¹E. B. Wilson, "Microwave spectroscopy in chemistry," *Science* **162**, 59–66 (1968).
- ²M. C. McCarthy and J. Gauss, "Exotic SiO₂H₂ isomers: Theory and experiment working in harmony," *The Journal of Physical Chemistry Letters* **7**, 1895–1900 (2016).
- ³M. Melosso, B. A. McGuire, F. Tamassia, C. Degli Esposti, and L. Dore, "Astronomical search of vinyl alcohol assisted by submillimeter spectroscopy," *ACS Earth and Space Chemistry* **3**, 1189–1195 (2019).
- ⁴F. C. De Lucia and P. A. Helminger, "Millimeterwave spectroscopy of active laser plasmas; the excited vibrational states of HCN," *The Journal of Chemical Physics* **67**, 4262–4267 (1977).
- ⁵K. Tanaka, M. Toshimitsu, K. Harada, and T. Tanaka, "Determination of the proton tunneling splitting of the vinyl radical in the ground state by millimeter-wave spectroscopy combined with supersonic jet expansion and ultraviolet photolysis," *The Journal of Chemical Physics* **120**, 3604–3618 (2004).
- ⁶G. G. Brown, B. C. Dian, K. O. Douglass, S. M. Geyer, S. T. Shipman, and B. H. Pate, "A broadband Fourier transform microwave spectrometer based on chirped pulse excitation," *Review of Scientific Instruments* **79**, 053103 (2008).
- ⁷G. B. Park and R. W. Field, "Perspective: The first ten years of broadband chirped pulse Fourier transform microwave spectroscopy," *The Journal of Chemical Physics* **144**, 200901 (2016).
- ⁸B. C. Dian, G. G. Brown, K. O. Douglass, and B. H. Pate, "Measuring picosecond isomerization kinetics via broadband microwave spectroscopy," *Science* **320**, 924–928 (2008).
- ⁹C. C. Womack, M.-A. Martin-Drumel, G. G. Brown, R. W. Field, and M. C. McCarthy, "Observation of the simplest criegee intermediate CH₂OO in the gas-phase ozonolysis of ethylene," *Science Advances* **1**, e1400105 (2015).
- ¹⁰D. P. Zaleski, R. Sivaramakrishnan, H. R. Weller, N. A. Seifert, B. Bross, D. H. amd Ruscic, K. B. Moore III, S. N. Elliott, A. V. Copan, L. B. Harding, S. J. Klippenstein, R. W. Field, and K. Prozument, "Substitution reactions in the pyrolysis of acetone revealed through a modeling, experiment, theory paradigm," *J. Am. Chem. Soc.* **143**, 3124–3142 (2021).

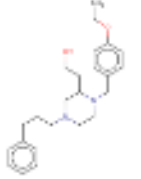
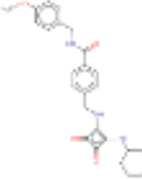
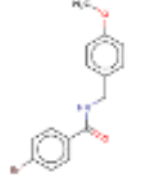
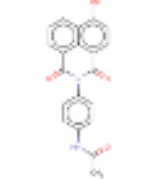
Category	Molecule 1	Molecule 2	Category	Molecule 1	Molecule 2
Lewis Structure			Lewis Structure		
R, MHz	246.49	246.72	R, MHz	660.44	660.58
A, MHz	229.68	229.84	A, MHz	649.67	649.58
B, MHz	67.43	67.64	B, MHz	87.64	87.48
C, MHz	58.81	58.90	C, MHz	80.16	80.82
μ_A^2, D^2 (ratio)	0.62 (1.00)	85.78 (1.00)	μ_A^2, D^2 (ratio)	20.31 (1.00)	18.62 (1.00)
μ_B^2, D^2 (ratio)	0.25 (0.40)	34.80 (0.41)	μ_B^2, D^2 (ratio)	1.92 (0.09)	1.64 (0.09)
μ_C^2, D^2 (ratio)	0.61 (0.98)	76.35 (0.89)	μ_C^2, D^2 (ratio)	13.73 (0.68)	12.77 (0.69)
Mol. Formula	C ₂₄ H ₃₄ N ₂ O ₂	C ₂₆ H ₂₉ N ₃ O ₄	Mol. Formula	C ₁₅ H ₂₀ BrNO ₂	C ₂₀ H ₁₄ N ₂ O ₄

FIG. 10. Two pairs of molecule conformers from the GEOM-Drug dataset that are twins. The rotational constants, squared dipole constants, squared dipole ratios (relative to the largest dipole), and molecular formulae are listed for each. For the pair of molecules on the left, their SMILES are CCOC1=CC=C(C=C1)CN2CCN(CC2CCO)CCCC3=CC=CC=C3 and O(C1=CC=C(CNC(C2C=CC(CNC3C(NC4CCCC4)=C(=O)C=3=O)=CC=2)=O)C=C1)C, respectively. For the pair of molecules on the right, their SMILES are O(C1CCC(CNC(C2C=CC(Br)=CC=2)=O)CC1)C and CC(=O)NC1=CC=C(C=C1)N2C(=O)C3=C4C(=C(C=C3)O)C=CC=C4C2=O, respectively.

- ¹¹Q. Borengasser, T. Hager, A. Kanaherarachchi, D. Troya, and B. M. Broderick, "Conformer-specific desorption in propanol ices probed by chirped-pulse millimeter-wave rotational spectroscopy," *J. Phys. Chem. Lett.* **14**, 6550–6555 (2023).
- ¹²S. Thawoos, N. Suas-David, R. M. Gurusinghe, M. Edlin, A. Behzadfar, J. Lang, and A. G. Suits, "Low temperature reaction kinetics inside an extended Laval nozzle: REMPI characterization and detection by broadband rotational spectroscopy," *The Journal of Chemical Physics* **159**, 214201 (2023).
- ¹³T. Guillaume, B. M. Hays, D. Gupta, I. R. Cooke, O. Abdelkader Khedaoui, T. S. Hearne, M. Drissi, and I. R. Sims, "Product-specific reaction kinetics in continuous uniform supersonic flows probed by chirped-pulse microwave spectroscopy," *The Journal of Chemical Physics* **160**, 204201 (2024).
- ¹⁴A. A. Byars, K. R. Kompally, E. Mechnick, R. E. Sonstrom, A. Mikhonin, J. L. Neill, R. Boetzel, J. MacGregor, C. Talicska, J. Li, *et al.*, "An automated, highly selective reaction monitoring instrument using molecular rotational resonance spectroscopy," *Precision Chemistry* (2024).
- ¹⁵M. C. McCarthy, K. L. K. Lee, P. B. Carroll, J. P. Porterfield, P. B. Changala, J. H. Thorpe, and J. F. Stanton, "Exhaustive product analysis of three benzene discharges by microwave spectroscopy," *The Journal of Physical Chemistry A* **124**, 5170–5181 (2020).
- ¹⁶J. L. Neill, L. Evangelisti, and B. H. Pate, "Analysis of isomeric mixtures by molecular rotational resonance spectroscopy," *Analytical Science Advances* **4**, 204–219 (2023).
- ¹⁷T. M. Wilson, A. Davis, R. E. Sonstrom, J. L. Neill, E. A. Ziebarth, A. Poulson, and R. E. Carlson, "Essential oil composition and enantioselective profile of *Agastache urticifolia* (Lamiaceae) and *Monardella odoratissima* (Lamiaceae) from Utah," *Molecules* **28**, 2249 (2023).
- ¹⁸J. A. Smith, K. B. Wilson, R. E. Sonstrom, P. J. Kelleher, K. D. Welch, E. K. Pert, K. S. Westendorff, D. A. Dickie, X. Wang, B. H. Pate, *et al.*, "Preparation of cyclohexene isotopologues and stereoisotopomers from benzene," *Nature* **581**, 288–293 (2020).
- ¹⁹D. W. Armstrong, M. Talebi, N. Thakur, M. F. Wahab, A. V. Mikhonin, M. T. Muckle, and J. L. Neill, "A gas chromatography-molecular rotational resonance spectroscopy based system of singular specificity," *Angewandte Chemie* **132**, 198–202 (2020).
- ²⁰M. D. Mills, R. E. Sonstrom, Z. P. Vang, J. L. Neill, H. N. Scolati, C. T. West, B. H. Pate, and J. R. Clark, "Enantioselective synthesis of enantioisotopomers with quantitative chiral analysis by chiral tag rotational spectroscopy," *Angewandte Chemie International Edition* **61**, e202207275 (2022).
- ²¹J. L. Neill, A. V. Mikhonin, T. Chen, R. E. Sonstrom, and B. H. Pate, "Rapid quantification of isomeric and dehalogenated impurities in pharmaceutical raw materials using mrr spectroscopy," *Journal of Pharmaceutical and Biomedical Analysis* **189**, 113474 (2020).
- ²²D. J. De Waele, S. Luyten, R. E. Sonstrom, J. Bogaerts, J. L. Neill, P. Viereck, K. Goossens, M. Baeten, N. Vervoort, and W. Herrebout, "Absolute configuration assignment of highly fluorinated carboxylic acids via VCD and MRR spectroscopy," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*

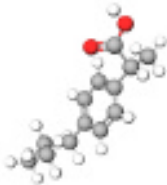
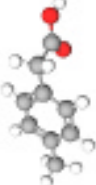
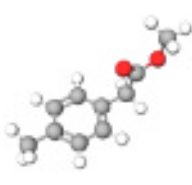
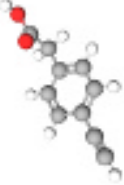
Category	Ibuprofen	Molecule 1	Molecule 2	Molecule 3
Lewis Structure				
R, MHz	4847.57	4140.35	3926.68	3927.83
A, MHz	4783.36	4050.26	3871.84	3877.05
B, MHz	587.25	613.18	464.52	447.15
C, MHz	523.04	601.60	460.29	443.15
μ_A^2, D^2 (ratio)	11.27 (1.00)	24.22 (1.00)	0.12 (0.03)	29.92 (1.00)
μ_B^2, D^2 (ratio)	7.27 (0.65)	<0.01 (<0.01)	0.06 (0.01)	0.04 (<0.01)
μ_C^2, D^2 (ratio)	<0.01 (<0.01)	4.11 (0.17)	4.70 (1.00)	3.56 (0.14)
Mol. Formula	$C_{13}H_{18}O_2$	$C_9H_{10}O_2$	$C_{10}H_{12}O_2$	$C_{10}H_8O_2$

FIG. 11. Four molecule conformers, including Ibuprofen from the NearIB dataset. These Molecules 1-3 are among the closest spectra to Ibuprofen in the NearIB dataset, but are not twins. The rotational constants, squared dipole constants, squared dipole ratios (relative to the largest dipole), and molecular formulae are listed for each. For all four molecules, the SMILES strings are CC(C)CC1=CC=C(C=C1)C(C)C(=O)O, CC1=CC=C(C=C1)CC(=O)O, COC(=O)Cc1ccc(C)cc1, and C#Cc1ccc(CC(=O)O)cc1 respectively.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0212632

- 306**, 123625 (2024).
- ²³Z. P. Vang, A. Reyes, R. E. Sonstrom, M. S. Holdren, S. E. Sloane, I. Y. Alansari, J. L. Neill, B. H. Pate, and J. R. Clark, "Copper-catalyzed transfer hydrodeuteration of aryl alkenes with quantitative isotopomer purity analysis by molecular rotational resonance spectroscopy," *Journal of the American Chemical Society* **143**, 7707–7718 (2021).
- ²⁴C. E. Rowell, J. Kaur, Y. Zhang, K. Z. Alzarini, R. O. Anyaeche, H. M. Ma, A. M. Little, T. Sharma, M. D. Nussbaum, D. A. Herschberger, *et al.*, "Application of laser-induced acoustic desorption for molecular rotational resonance spectroscopy," *Analytical Chemistry* **96**, 6255–6263 (2024).
- ²⁵V. Barone, "Quantum chemistry meets high-resolution spectroscopy for characterizing the molecular bricks of life in the gas-phase," *Physical Chemistry Chemical Physics* **26**, 5802–5821 (2024).
- ²⁶V. Barone and C. Puzzarini, "Gas-phase computational spectroscopy: The challenge of the molecular bricks of life," *Annual Review of Physical Chemistry* **74**, 29–52 (2023).
- ²⁷M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme, "Best-practice DFT protocols for basic molecular computational chemistry," *Angewandte Chemie International Edition* **61**, e202205735 (2022).
- ²⁸K. L. K. Lee and M. McCarthy, "Bayesian analysis of theoretical rotational constants from low-cost electronic structure methods," *The Journal of Physical Chemistry A* **124**, 898–910 (2020).
- ²⁹N. A. Seifert, M. J. Davis, and K. Prozument, "Prediction of molecular structures from rotational constants: A proposal for solving the inverse problem," *International Symposium on Molecular Spectroscopy* (2020), 10.15278/isms.2020.RI07.
- ³⁰M. McCarthy and K. L. K. Lee, "Molecule identification with rotational spectroscopy and probabilistic deep learning," *The Journal of Physical Chemistry A* **124**, 3002–3017 (2020).
- ³¹A. H. Cheng, A. Lo, S. Miret, B. H. Pate, and A. Aspuru-Guzik, "Determining 3D structure from molecular formula and isotopologue rotational spectra in natural abundance with reflection-equivariant diffusion," *The Journal of Chemical Physics* **160**, 124115 (2024).
- ³²A. Tarantola, *Inverse problem theory and methods for model parameter estimation* (SIAM, 2005).
- ³³M. Kac, "Can one hear the shape of a drum?" *The American Mathematical Monthly* **73**, 1–23 (1966).
- ³⁴C. Gordon, D. L. Webb, and S. Wolpert, "One cannot hear the shape of a drum," *Bulletin of the American Mathematical Society* **27**, 134–138 (1992).
- ³⁵M. Bertero, P. Boccacci, and C. De Mol, *Introduction to inverse problems in imaging* (CRC press, 2021).
- ³⁶I. Dokmanić, Y. M. Lu, and M. Vetterli, "Can one hear the shape of a room: The 2-D polygonal case," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, 2011) pp. 321–324.
- ³⁷S. Park, I. Lee, J. Kim, N. Park, and S. Yu, "Hearing the shape of a drum for light: Isospectrality in photonics," *Nanophotonics* **11**, 2763–2778 (2021).
- ³⁸D. Pursey, "Isometric operators, isospectral Hamiltonians, and supersymmetric quantum mechanics," *Physical Review D* **33**, 2267 (1986).
- ³⁹E. Heilbronner and T. B. Jones, "Spectral differences between "isospectral" molecules," *Journal of the American Chemical Society* **100**, 6506–6507 (1978).
- ⁴⁰C. A. Coulson, B. O'Leary, and R. B. Mallion, "Hückel theory for organic chemists," (No Title) (1978).
- ⁴¹H. H. Günthard and H. Primas, "Zusammenhang von Graphentheorie und MO-Theorie von Molekeln mit Systemen konjugierter Bindungen," *Helvetica Chimica Acta* **39**, 1645–1653 (1956).
- ⁴²L. Von Collatz and U. Sinogowitz, "Spektren endlicher grafen: Wilhelm Blaschke zum 70. Geburtstag gewidmet," in *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, Vol. 21 (Springer, 1957) pp. 63–77.
- ⁴³J. R. Dias, "Almost-isospectral conjugated polyene molecules," *Chemical physics letters* **253**, 305–312 (1996).
- ⁴⁴W. Herndon and M. Ellzey Jr, "Isospectral graphs and molecules," *Tetrahedron* **31**, 99–107 (1975).
- ⁴⁵E. R. Van Dam and W. H. Haemers, "Which graphs are determined by their spectrum?" *Linear Algebra and its applications* **373**, 241–272 (2003).
- ⁴⁶H. Hosoya, "Chemistry-relevant isospectral graphs. acyclic conjugated polyenes," *Croatica Chemica Acta* **89**, 455–461 (2016).
- ⁴⁷J. Schrier, "Can one hear the shape of a molecule (from its Coulomb matrix eigenvalues)?" *Journal of Chemical Information and Modeling* **60**, 3804–3811 (2020).
- ⁴⁸M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. Von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," *Physical Review Letters* **108**, 058301 (2012).
- ⁴⁹H. W. Kroto, *Molecular rotation spectra* (John Wiley & Sons, 1975).
- ⁵⁰K. Hellwege and L. C. Green, "Landolt-Börnstein, Numerical data and functional relationships in science and technology," *American Journal of Physics* **35**, 291–292 (1967).
- ⁵¹R. J. Silbey and J. L. Kinsey, "On the preponderance of near-prolate rotors among polyatomic molecules," *The Journal of chemical physics* **88**, 4100–4100 (1988).
- ⁵²H. M. Pickett, "The fitting and prediction of vibration-rotation spectra with spin interactions," *Journal of Molecular Spectroscopy* **148**, 371–377 (1991).
- ⁵³C. M. Western, "PGOPHER: A program for simulating rotational, vibrational and electronic spectra," *Journal of Quantitative Spectroscopy and Radiative Transfer* **186**, 221–242 (2017).
- ⁵⁴N. A. Seifert, I. A. Finneran, C. Perez, D. P. Zaleski, J. L. Neill, A. L. Steber, R. D. Suenram, A. Lesarri, S. T. Shipman, and B. H. Pate, "AUTOFIT, an automated fitting tool for broadband rotational spectra, and applications to 1-hexanal," *Journal of Molecular Spectroscopy* **312**, 13–21 (2015).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0212632

- ⁵⁵G. Di Modica, L. Evangelisti, L. Foschini, A. Maris, and S. Melandri, "Testing the scalability of the HS-AUTOFIT tool in a high-performance computing environment," *Electronics* **10**, 2251 (2021).
- ⁵⁶L. Yeh, L. Satterthwaite, and D. Patterson, "Automated, context-free assignment of asymmetric rotor microwave spectra," *The Journal of chemical physics* **150** (2019).
- ⁵⁷S. Cooke, P. Ohring, *et al.*, "Decoding pure rotational molecular spectra for asymmetric molecules," *Journal of Spectroscopy* **2013** (2013).
- ⁵⁸D. P. Zaleski and K. Prozument, "Automated assignment of rotational spectra using artificial neural networks," *The Journal of Chemical Physics* **149**, 104106 (2018).
- ⁵⁹N. A. Seifert, K. Prozument, and M. J. Davis, "Computational optimal transport for molecular spectra: The fully discrete case," *The Journal of Chemical Physics* **155** (2021).
- ⁶⁰N. A. Seifert, K. Prozument, and M. J. Davis, "Computational optimal transport for molecular spectra: The semi-discrete case," *The Journal of Chemical Physics* **156** (2022).
- ⁶¹N. A. Seifert, K. Prozument, and M. J. Davis, "Computational optimal transport for molecular spectra: The fully continuous case," *The Journal of Chemical Physics* **159** (2023).
- ⁶²P. B. Carroll, K. L. K. Lee, and M. C. McCarthy, "A high speed fitting program for rotational spectroscopy," *Journal of Molecular Spectroscopy* **379**, 111467 (2021).
- ⁶³J. Kraitchman, "Determination of molecular structure from microwave spectroscopic data," *American Journal of Physics* **21**, 17–24 (1953).
- ⁶⁴C. Pérez, S. Lobsiger, N. A. Seifert, D. P. Zaleski, B. Temelso, G. C. Shields, Z. Kisiel, and B. H. Pate, "Broadband Fourier transform rotational spectroscopy for structure determination: The water heptamer," *Chemical Physics Letters* **571**, 1–15 (2013).
- ⁶⁵C. Puzzarini, "Rotational spectroscopy meets theory," *Physical Chemistry Chemical Physics* **15**, 6595–6607 (2013).
- ⁶⁶A. P. Bento, A. Hersey, E. Félix, G. Landrum, A. Gaulton, F. Atkinson, L. J. Bellis, M. De Veij, and A. R. Leach, "An open source chemical structure curation pipeline using RDKit," *Journal of Cheminformatics* **12**, 1–16 (2020).
- ⁶⁷N. M. O'Boyle, "Towards a universal SMILES representation-A standard method to generate canonical SMILES based on the InChI," *Journal of Cheminformatics* **4**, 1–14 (2012).
- ⁶⁸A. G. Marshall and C. L. Hendrickson, "High-resolution mass spectrometers," *Annu. Rev. Anal. Chem.* **1**, 579–599 (2008).
- ⁶⁹R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. Von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," *Scientific Data* **1**, 1–7 (2014).
- ⁷⁰J. Hoja, L. Medrano Sandonas, B. G. Ernst, A. Vazquez-Mayagoitia, R. A. DiStasio Jr, and A. Tkatchenko, "QM7-X, a comprehensive dataset of quantum-mechanical properties spanning the chemical space of small organic molecules," *Scientific Data* **8**, 43 (2021).
- ⁷¹S. Axelrod and R. Gomez-Bombarelli, "GEOM, energy-annotated molecular conformations for property prediction and molecular generation," *Scientific Data* **9**, 185 (2022).
- ⁷²P. Pracht, F. Bohle, and S. Grimme, "Automated exploration of the low-energy chemical space with fast quantum chemical methods," *Physical Chemistry Chemical Physics* **22**, 7169–7192 (2020).
- ⁷³S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, *et al.*, "PubChem 2023 update," *Nucleic Acids Research* **51**, D1373–D1380 (2023).
- ⁷⁴N. O. Lüttschwager, T. N. Wassermann, R. A. Mata, and M. A. Suhm, "The last globally stable extended alkane," *Angewandte Chemie International Edition* **52**, 463–466 (2013).
- ⁷⁵K. D. Hensel, C. Styger, W. Jäger, A. Merer, and M. Gerry, "Microwave spectra of metal chlorides produced using laser ablation," *The Journal of Chemical Physics* **99**, 3320–3328 (1993).
- ⁷⁶A. Lesarri, S. Mata, J. C. López, and J. L. Alonso, "A laser-ablation molecular-beam fourier-transform microwave spectrometer: The rotational spectrum of organic solids," *Review of Scientific Instruments* **74**, 4799–4804 (2003).
- ⁷⁷C. Pérez, M. T. Muckle, D. P. Zaleski, N. A. Seifert, B. Temelso, G. C. Shields, Z. Kisiel, and B. H. Pate, "Structures of cage, prism, and book isomers of water hexamer from broadband rotational spectroscopy," *Science* **336**, 897–901 (2012).