frontiers in Chemistry

REVIEW
published: 07 August 2019
doi: 10.3389/fchem.2019.00519

Augmenting Basin-Hopping With Techniques From Unsupervised Machine Learning: Applications in Spectroscopy and Ion Mobility

*Ce Zhou, Christian Ieritano and William Scott Hopkins**

Department of Chemistry, University of Waterloo, Waterloo, ON, Canada

Evolutionary algorithms such as the basin-hopping (BH) algorithm have proven to be useful for difficult non-linear optimization problems with multiple modalities and variables. Applications of these algorithms range from characterization of molecular states in statistical physics and molecular biology to geometric packing problems. A key feature of BH is the fact that one can generate a coarse-grained mapping of a potential energy surface (PES) in terms of local minima. These results can then be utilized to gain insights into molecular dynamics and thermodynamic properties. Here we describe how one can employ concepts from unsupervised machine learning to augment BH PES searches to more efficiently identify local minima and the transition states connecting them. Specifically, we introduce the concepts of similarity indices, hierarchical clustering, and multidimensional scaling to the BH methodology. These same machine learning techniques can be used as tools for interpreting and rationalizing experimental results from spectroscopic and ion mobility investigations (e.g., spectral assignment, dynamic collision cross sections). We exemplify this in two case studies: (1) assigning the infrared multiple photon dissociation spectrum of the protonated serine dimer and (2) determining the temperature-dependent collision cross-section of protonated alanine tripeptide.

Keywords: serine dimer, polyalanine, collision cross section, IRMPD, hierarchical clustering, potential energy surface, global optimization, vibrational spectroscopy

OPEN ACCESS

Edited by:
Emilio Martinez-Nunez,
University of Santiago de Compostela, Spain

Reviewed by:
Daniel Pelaez,
Université de Lille, France
Tomás González-Lezana,
Spanish National Research Council (CSIC), Spain

****Correspondence:***
William Scott Hopkins
scott.hopkins@uwaterloo.ca

Specialty section:
This article was submitted to Theoretical and Computational Chemistry,
a section of the journal Frontiers in Chemistry

Received: *30 April 2019*
Accepted: *08 July 2019*
Published: *07 August 2019*

Citation:
Zhou C, Ieritano C and Hopkins WS (2019) Augmenting Basin-Hopping With Techniques From Unsupervised Machine Learning: Applications in Spectroscopy and Ion Mobility. Front. Chem. 7:519. doi: 10.3389/fchem.2019.00519

INTRODUCTION

Molecular global optimization (GO) to identify the chemically-relevant species on hypergeometric potential energy surfaces (PESs) provides both rationalizations and predictions of experimental observations by relating thermodynamic and kinetic properties to the accessible local minima and the transition states (TSs) that connect them (Scheraga, 1992; Piela et al., 1994; Wales and Doye, 1997; Wales and Scheraga, 1999). Basin-hopping (BH) is a technique for GO that is based on the iterative approach of performing random perturbation of geometric coordinates, local optimization of a model potential energy function, and accepting or rejecting the perturbed coordinates based on the value of the minimized function (Wales and Doye, 1997; Wales et al., 1998; Wales and Scheraga, 1999; Lecours et al., 2014). Use of the BH algorithm for searching molecular PESs was outlined by Wales and Doye in their 1997 article "Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms," (Wales and Doye, 1997) which describes how the technique transforms the PES into a collection of interpenetrating staircases wherein each stair/plateau on the transformed surface is associated with a stationary point (usually local minimum) of the original potential energy landscape. **Figure 1** shows a flow diagram outlining the general procedure of the BH search algorithm. The key feature of the BH algorithm is the inclusion of assessment criteria for accepting or rejecting a newly distorted input geometry. One of these criteria is the definite replacement of the lowest energy structure identified by the BH routine with the currently optimized structure if that structure has a lower energy. A second key criterion is a conditional acceptance of the distorted geometry by assessing the statistical accessibility of the optimized structure based on a pre-defined energy window. For example, one can define a Boltzmann distribution at a given temperature with respect to the current lowest energy structure and assess the probability of accessing the newly generated stationary point. Thus, the BH algorithm has a bias toward low energy structures and is a good option for identifying the global minimum (GM) and local minima that may be present in an ensemble under thermal equilibrium conditions.

To further improve the efficiency of a BH search, one can include additional criteria for assessment of distorted molecular geometries prior to optimization. For example, one might choose to reject structures in which inter-atomic distances are less than some pre-defined threshold, or one might choose to define an interaction volume to prevent molecular/cluster dissociation (Lecours et al., 2014). It is also common to select specific degrees of freedom (DoFs) for random distortion while freezing others; one might choose to search the conformational space defined by molecular dihedral angles while leaving the distances between chemically bonded atoms fixed (Hopkins et al., 2013, 2015). There are several other works which employ more dramatic changes to the underlying BH algorithm. For example, Leary proposed a version in which only the replacement criterion is employed in the evaluation (i.e., no statistically accessible energy window is specified) (Leary, 2000). In other works, Röder and Wales propose a mutational BH algorithm to optimize biomolecules (Röder and Wales, 2018), and Kim et al. combine BH with Coulomb matrix analysis to sample reaction intermediates (Kim et al., 2014). While these variants have all been successful in the task at hand, the fact that the basic BH algorithm often requires tailoring highlights the inherent drawbacks in the BH methodology.

One principal short-coming of the BH algorithm that practitioners must be aware of is that the method is not deterministic; i.e., identifying the GM via a finite, stochastic search is not guaranteed. Confidence in BH search results come from a satisfactory agreement with experimental observations and/or the consistency of results from several parallel simulations with different initial conditions. A second potential short-coming is the fact that, due to performance considerations, BH calculations are often conducted with relatively low-level model chemistries (e.g., molecular mechanics), which may not be accurate enough for certain molecular systems. Finally, practitioners must be aware that a BH search may be kinetically trapped in a local potential minimum if the thermal energy (*viz*. temperature) of the simulation is set too low. In fact, in some cases BH searches of PESs are non-ergodic regardless of simulation temperature. For example, consider the case of protonated *para*-aminobenzoic acid, which can exhibit protonation on either the carbonyl oxygen atom or the amine nitrogen atom in the gas phase (Tian and Kass, 2009; Schmidt et al., 2011; Campbell et al., 2012, 2016). If one were to assume that the protonation site of *para*-aminobenzoic acid were the nitrogen center (as is the case in protic solution) and model the system as a molecular cation using a molecular mechanics force field, the O-protonated isomer (which is the gas phase global minimum) would not be identified without modifying the atomic connectivity during the BH search (Tian and Kass, 2008; Campbell et al., 2012, 2016). To overcome this systematic limitation, one must treat the charge-carrying proton as a separate moiety in the simulation and/or augment the BH framework with the chemical intuition of the user (i.e., manually identify both prototropic isomers and conduct BH searches for each of them).

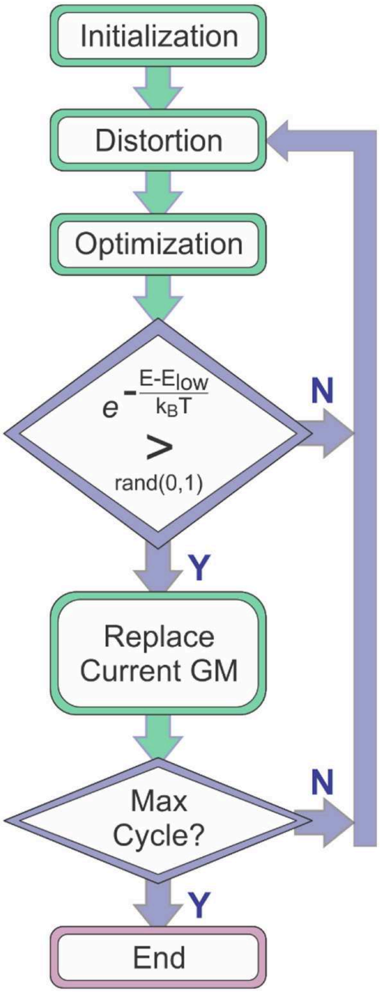


FIGURE 1 | The general procedure of the basin-hopping algorithm. E_{low} is the energy of the lowest energy species identified to that point in the search (i.e., the current global minimum, GM).

Here, we describe how the basin-hopping algorithm can be employed to reliably model gas phase cluster and molecular systems for comparison with observations from spectroscopy and ion mobility experiments. To model our experimental observations, we require theoretical predictions from a collection of local minima, which do not necessarily include the global minimum, and an efficient method to find matches between the predictions and the observations. In some cases, it is also desirable to identify the TSs that connect minima to assess thermodynamic accessibility of the various isomers / conformers. These two requirements present two notable challenges for the BH methodology. The first challenge, related to the principal short-coming mentioned above, is the necessity to accurately track the explored regions of the PES. In doing so, one not only identifies a set of local minima, but also gains useful information for directing the BH search toward regions of the PES that are relatively unexplored. The second challenge is the accurate and efficient identification of the TSs that connect local minima. To overcome these challenges, we collect the nuclear configuration data that is generated during the BH search and utilize this data as described in Section Augmenting the BH Algorithm. Specifically, in Section Assessing Geometric Similarity we describe how one can utilize similarity functions and hierarchical clustering, which are concepts generally associated with unsupervised machine learning, to assess the uniqueness of the local minima and guide PES searches. We then discuss the interpolation of geometries to identify intermediate local minima and to create guess geometries for TS searches in Section Interpolating Intermediate Geometries. In Section Application of BH Search Results, we outline our methods for employing our BH results to assign the spectral carriers (Section Case Study 1: The IR Spectrum of the Protonated Serine Dimer) and to model temperature-dependent structures (Section Case Study 2: Dynamic Collision Cross Section of Protonated Alanine Tripeptide) of geometrically-fluxional species. Finally, we summarize our perspective and highlight open questions in Section Conclusions.

AUGMENTING THE BH ALGORITHM

As mentioned in section Introduction, several variations to the BH algorithm have been proposed to address specific challenges in searching complex potential energy landscapes (Leary, 2000; Kim et al., 2014; Röder and Wales, 2018). For our purposes, where it is necessary to identify a collection of local minima that are representative of the species present in experimental ensembles, we require a faithful mapping of the molecular PES. To improve the efficiency and PES coverage of the BH algorithm, we introduce a method of comparing the geometries of local minima. This comparison, which is derived from a similarity function, provides a more rigorous identification of unique isomeric species and insight into which regions of the PES may require additional exploration.

In analogy to the spatial distance between two locations on a map, a similarity function quantifies the similarity of two conformations, A and B, in conformation space. The function, usually denoted as d(A,B), is non-negative ($d\left(A,B\right) \geq 0$), symmetric ($d\left(A,B\right) = d(B,A)$) and has zero value only when two identical elements are evaluated ($d\left(A,A\right) = 0$) (Locatelli and Schoen, 2013). The similarity function can be used in one of three ways: qualification, quantification, and interpolation. Qualification usage implies that the function need only tell if two input structures are identical. Quantification usage provides a metric for how much difference is there between two structures; for example, is structure A more similar to structure B than to structure C? Interpolation usage means that, given two structures, A and B, and an arbitrary interpolation factor, $\lambda \in (0,1)$, there exist one or more structures, C, satisfying:

$$d\left(A,B\right) = \frac{1}{\lambda} d\left(A,C\right) = \frac{1}{1-\lambda} d\left(B,C\right) \tag{1}$$

If the function d satisfies triangular inequality $d\left(A,B\right) + d\left(B,C\right) \geq d(A,C)$), the structure C is unique, and d is a metric of the conformation space (Choudhary, 2003). Note that special treatment is required if A and B have different numbers of atoms (i.e., if A and B are of different dimension); this tends not to be the case in simulations of chemical systems. The interpolation mechanism is of central importance not only to a number of GO algorithms, such as particle swarm optimization (Eberhart and Yuhui, 2001), differential evolution (Storn and Price, 1997), and DIRECT (Jones et al., 1993), but also to unsupervised machine learning techniques such as the self-organizing map (Kohonen, 1990) and the growing neural gas (Martinez and Schulten, 1991; Fritzke, 1994). In qualitative comparisons, the similarity function need only account for the translational, rotational, and permutational invariance under a given molecular representation; structural equivalence only occurs between species of identical composition. Such invariance properties are either embedded in the mathematical definition of the molecular representation or they are achieved via manually aligning the two molecular systems prior to evaluating their similarity. Examples of such representations include the conventional skeletal chemical formula and the SMILEs code used in compound database systems (Weininger, 1988; Rahman et al., 2009; Heller et al., 2013). In quantitative comparisons, the similarity of two structures is specified by a real number. These similarity indices are useful in discriminating visited regions of the PES (e.g., well-sampled vs. poorly-sampled regions), which can be assessed using unsupervised machine learning analyses like hierarchical clustering and multidimensional scaling (MDS) (Wickelmaier, 2003; Borg and Groenen, 2005). Most similarity functions used for quantitation purposes are defined by the normal (e.g., the root-mean-square deviation of atomic positions, RMSD) (Kabsch, 1976) or reciprocal (e.g., the Coulomb matrix) (Montavon et al., 2012) interatomic distances, although electron density-based similarity functions have found use in drug discovery (Cereto, 2015; Kumar and Zhang, 2018). To implement structural interpolation, the back conversion from desired similarity constraints to a concrete structure is required. This technique enables generation of intermediate geometries for TS calculations (e.g., QST3) (Peng and Bernhard Schlegel, 1993; Peng et al., 1996), and it can also be used to guide BH searches of specified regions of the PES along isomerization pathways between two isomers. Furthermore, by implementing structural interpolation, one creates the opportunity to incorporate other GO techniques (e.g., particle swarm optimization) (Kennedy and Eberhart, 1995; Call et al., 2007; Shi et al., 2019) and machine learning techniques (e.g., growing neural gas) (Martinez and Schulten, 1991; Fritzke, 1994) into the BH algorithm. In practice, rather than an explicit analytical approach, structure interpolation can be achieved implicitly via local optimizations with a tolerable loss of accuracy. In our research, to efficiently use the nuclear configuration information from the BH simulation, we introduce both Euclidean distance matrix-based and cosine distance-based similarity functions together with the necessary techniques to accomplish structural interpolation. The mathematical and implementation details are described below.

Assessing Geometric Similarity

To begin assessing the similarity between two molecular geometries, one must first select an appropriate similarity function. One option, the Euclidean distance matrix representation (***D***) of a molecule, is simply the collection of all interatomic distances as per (Gentle, 2007):

$$\mathbf{D}_{ij} = |\vec{r_i} - \vec{r_j}| \tag{2}$$

where $\vec{r_i}$ and $\vec{r_j}$ are the positional vectors (in Cartesian coordinates) of atoms i and j. Within the distance matrix representation, the similarity function is defined as the sum of the absolute difference between each atom pair for structures A and B:

$$d\left(\mathbf{D}_A, \mathbf{D}_B\right) = \sum i, j > i|\mathbf{D}_{A,ij} - \mathbf{D}_{B,ij}| \tag{3}$$

The distance matrix is a symmetric matrix with diagonal elements of zero. This representation is translationally and rotationally invariant, but not permutationally invariant (*viz.* identical nuclei are not necessarily chemically equivalent). Thus, in practice, the atom labeling should be adjusted such that the similarity index (the value of the similarity function) of the two input molecules is minimized. It should be noted that the memory requirement of this representation scales quadratically with the number of atoms. Consequently, the distance matrix approach is not a good choice for dealing with very large systems.

A second option is to represent the molecular nuclear configuration as a vector, $\vec{R}$ (Fu and Hopkins, 2018), containing the mass-weighted distance between each atom and the molecular center-of-mass:

$$\vec{R}_{COM} = \frac{\sum_i^{m_i} \vec{r_i}}{\sum_i m_i} \tag{4}$$

$$\vec{R_i} = m_i|\vec{r_i} - \vec{R}_{COM}| \tag{5}$$

Where m_i and $\vec{r}_i$ are the mass and the distance to the center-of-mass for the i^{th} atom. Given that the mass-weighted distance vector representation is in the center-of-mass frame, one can then calculate the cosine distance between the vectors for isomers A and B as per:

$$d(\vec{R_A}, \vec{R_B}) = \frac{\cos^{-1}\left(s(\vec{R_A}, \vec{R_B})\right)}{\pi} \tag{6}$$

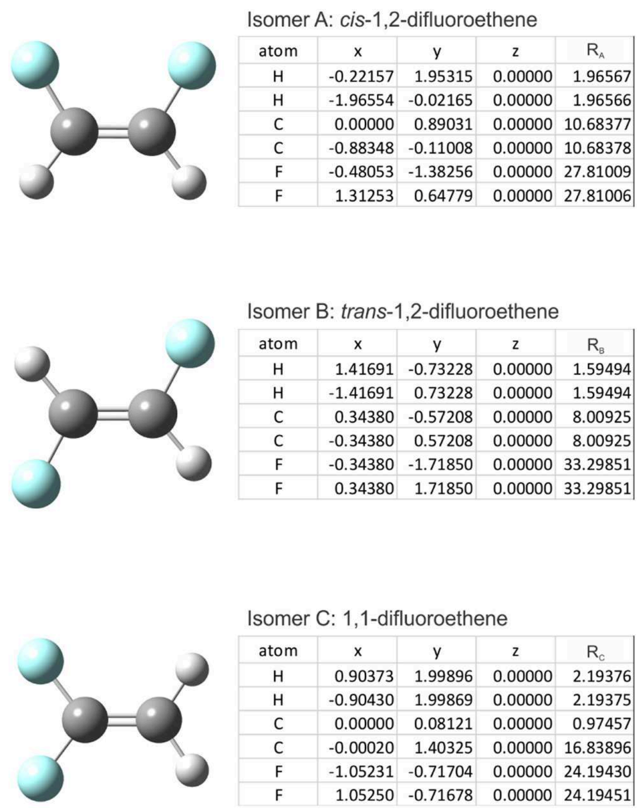
Where

$$s(\vec{R_A}, \vec{R_B}) = \frac{\vec{R_A} \cdot \vec{R_B}}{\left|\vec{R_A}\right|\left|\vec{R_B}\right|} \tag{7}$$

Again, this representation is translationally and rotationally invariant. However, care should be taken to ensure that the identity of the ith atom is retained throughout the BH search so that one compares the same atoms in each unique geometric structure. Alternatively, one might choose an operational convention whereby the resulting vector is sorted (e.g., smallest to largest values) prior to calculating cosine distance; this introduces a permutational invariance to the treatment for low symmetry systems. In contrast to the quadratic scaling of the distance matrix, the mass-weighted distance vector scales linearly with number of atoms. However, as a trade-off, the mass-weighted distance vector representation is less effective than the distance matrix approach in discriminating between conformers of highly symmetric species. For example, the mass-weighted distance vector representation is unable to distinguish square planar and tetrahedral conformations of methane given identical C–H bond length. Nevertheless, the uniqueness of the isomer-vector correspondence is still largely guaranteed in most cases in which only low symmetry structures are considered, particularly when relative energies are also considered in distinguishing isomeric/conformeric species.

The cosine similarity (Equation 7) ranges from −1 (meaning exactly opposite) to +1 (meaning identical). However, in practice, the cosine similarity for real molecular structures ranges from 0 to 1 since the center-of-mass vector is constructed from real space distances, which are always positive. Thus, two identical structures exhibit mass-weighted distance vectors with zero angular distance between them, and angular distances between vectors increase as the differences between the geometric structures of the associated isomers increase. For example, consider the isomers *cis*-1,2-difluoroethene, *trans*-1,2-difluoroethene, and 1,1-difluoroethene shown below in **Figure 2**. By inspection, one can identify that the mass-weighted distance vectors for the *cis*-1,2-difluoroethene and *trans*-1,2-difluoroethene isomers (R_A, R_B) are more like one another than they are to that of the 1,1-difluoroethene isomer (R_C). This is confirmed when calculating the cosine distances (see **Table 1**).

Calculating the distances between molecular structures facilitates analysis through agglomerative hierarchical clustering (Day and Edelsbrunner, 1984). This analysis provides a visual representation of the similarity of geometric structures—via production of a dendrogram plot—and therefore provides some insight into which species occupy similar regions of the potential energy landscape with respect to the mass-weighted nuclear coordinates. There are several methods available for analysis via agglomerative hierarchical clustering (Day and Edelsbrunner, 1984). One option for this analysis is the weighted pair group method with arithmetic mean (WPGMA), developed by Sokal and Michener (Michener and Sokal, 1957; Sokal and Michener, 1958). In each iteration of the WPGMA algorithm, the two nearest species (P and Q) are combined into a higher-level group P ∪ Q, thereby reducing the dimension of the $m \times m$ distance



Isomer A: *cis*-1,2-difluoroethene

atom	x	y	z	R_A
H	-0.22157	1.95315	0.00000	1.96567
H	-1.96554	-0.02165	0.00000	1.96566
C	0.00000	0.89031	0.00000	10.68377
C	-0.88348	-0.11008	0.00000	10.68378
F	-0.48053	-1.38256	0.00000	27.81009
F	1.31253	0.64779	0.00000	27.81006

Isomer B: *trans*-1,2-difluoroethene

atom	x	y	z	R_B
H	1.41691	-0.73228	0.00000	1.59494
H	-1.41691	0.73228	0.00000	1.59494
C	0.34380	-0.57208	0.00000	8.00925
C	-0.34380	0.57208	0.00000	8.00925
F	-0.34380	-1.71850	0.00000	33.29851
F	0.34380	1.71850	0.00000	33.29851

Isomer C: 1,1-difluoroethene

atom	x	y	z	R_C
H	0.90373	1.99896	0.00000	2.19376
H	-0.90430	1.99869	0.00000	2.19375
C	0.00000	0.08121	0.00000	0.97457
C	-0.00020	1.40325	0.00000	16.83896
F	-1.05231	-0.71704	0.00000	24.19430
F	1.05250	-0.71678	0.00000	24.19451

FIGURE 2 | The structures of **(top)** *cis*-1,2-difluoroethene, **(middle)** *trans*-1,2-difluoroethene, and **(bottom)** 1,1-difluoroethene. **(Inset Tables)** atomic coordinates and the mass-weighted distance vectors. Geometries were optimized at the PM6 level of theory as implemented in Gaussian 16 (Frisch et al., 2016).

TABLE 1 | The cosine distance matrix for *cis*-1,2-difluoroethene, *trans*-1,2-difluoroethene, and 1,1-difluoroethene.

Distance	*cis*-1,2-difluoro	*trans*-1,2-difluoro	1,1-difluoro
cis-1,2-difluoro	0	0.04200	0.09497
trans-1,2-difluoro	0.04200	0	0.10219
1,1-difluoro	0.09497	0.10219	0

Geometries were optimized at the PM6 level of theory as implemented in Gaussian 16 (Frisch et al., 2016).

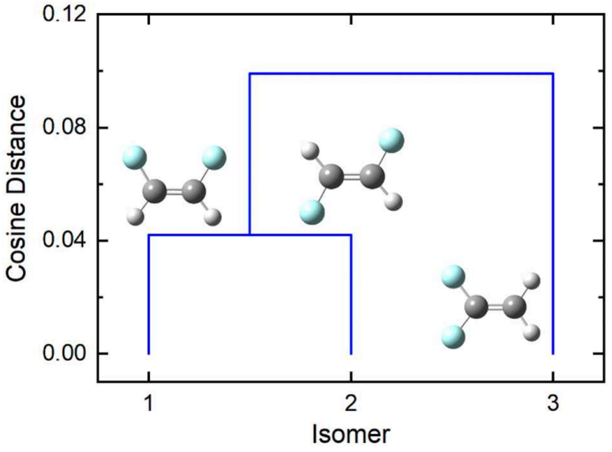


FIGURE 3 | The cosine distance dendrogram for difluoroethene. Molecular geometries were optimized at the PM6 level of theory as implemented in Gaussian 16 (Frisch et al., 2016).

matrix (e.g., **Table 1**) by one row and one column. The distance between group P ∪ Q and another group R is the arithmetic mean of the distances between the members of P ∪ Q and R, i.e.,:

$$d_{(P\cup Q),R} = \frac{d_{P,R} + d_{Q,R}}{2} \tag{8}$$

In the case of difluoroethene (**Figure 2** and **Table 1**), the smallest cosine distance of 0.042 between the *cis*- and *trans*-1,2-difluoroethene isomers would lead to their clustering as P ∪ Q, and the distance between this higher-level group and the 1,1-difluoroethene isomer would be (0.09497 + 0.10219)/2 = 0.09858. A dendrogram showing the hierarchical clustering of the isomers of difluoroethene is provided in **Figure 3**. By inspection of the dendrogram one can immediately see that the *cis*- and *trans*- isomers of 1,2-difluoroethene isomers are more closely related geometrically than either of these isomers is related to 1,1-difluoroethene.

Interpolating Intermediate Geometries

When searching complex PESs to find local minima or TSs, it is sometimes useful to interpolate geometries that are intermediate to two previously identified isomers. For example, consider the case in which a set of isomeric species has been identified, but one is very dissimilar from the others as determined by the geometric analysis described above. This might indicate that the BH search has become kinetically trapped and more attention should be paid to the region of the PES associated with the isolated structure. It is then useful to explore the PES between the more extensively mapped region and the region associated with the isolated structure to search for intermediates along the isomerization pathway and/or identify barriers to isomer interconversion. For the purpose of generating initial guess structures for the BH algorithm or for QST3 TS calculations, precise interpolation is not always necessary; (Peng and Bernhard Schlegel, 1993; Peng et al., 1996) most of the time interpolation can be accomplished implicitly, thereby improving the efficiency of the PES mapping. Currently, we have implemented two classes of implicit interpolation methods, one based on Monte Carlo sampling and the other based on molecular dynamics simulation.

Since the acceptance criteria are replaceable as a standard module in the evaluation part of the BH framework, instead of searching for low energy structures, one can choose to sample structures between two given minima on the PES within specified similarity constrains. Thus, a Monte Carlo with minimization approach can be established along a specified path/region of the PES. By applying an upper threshold to the distance of the sampled structure from the minima, one can constrain the search to a hyperdimensional ellipsoidal space between the two minima of interest. Within the distance matrix representation, the interpolation can also be accomplished with optimization on an interpolated artificial force field. Similar to the idea of the artificial force induced reaction (Maeda et al., 2014), the interpolated structure is obtained by minimizing a molecular mechanics-type force field, $\boldsymbol{V}$:

$$V(D_C) = \frac{\chi}{\bar{r}_{ij}} \left(D_{C,ij} - \bar{r}_{ij}\right)^2 \tag{9}$$

where χ is an arbitrary constant that facilitates optimization, and $D_{C,ij}$ and $\bar{r}_{ij}$, are the actual and expected interatomic distance of the interpolated structure. $\bar{r}_{ij}$ is constructed from the two minima, D_A and D_B and the interpolation factor, λ ($0 \leq \lambda \leq 1$) as per:

$$\bar{r}_{ij} = \lambda D_{A,ij} + (1-\lambda) D_{B,ij} \tag{10}$$

The force field is thus a collection of harmonic terms whose force constant is inversely proportional to $\bar{r}_{ij}$. Compared to the Monte Carlo approach, using this force field approach in conjunction with standard geometry optimization techniques is expected to be more efficient at identifying intermediate structures owing to the reduced and more pertinent search space.

APPLICATION OF BH SEARCH RESULTS

Experimental measurements are typically concerned with probing ensembles, rather than single molecules. Consequently, it is necessary to identify which structures are present in the probed ensemble and the relative populations of those species. This can be particularly challenging for chemical systems that are kinetically trapped in a relatively high-energy region of the PES and for systems that are fluxional (i.e., those that can easily access multiple minima on the experimental time scale). To demonstrate the potential of our augmentation to the original BH method, we describe our efforts to model the infrared multiple photon dissociation (IRMPD) spectrum of proton-bound serine dimer and the temperature-depending collision cross section (CCS) of protonated alanine tripeptide, $[\mathrm{AAA+H}]^+$.

Case Study 1: The IR Spectrum of the Protonated Serine Dimer

IRMPD spectroscopy has become one of the most effective techniques for determining the structure of molecular ions (Jašíková and Roithová, 2018). Ion spectra are recorded by isolating a specified *m/z* species in an ion trap and monitoring the fragmentation efficiency of the molecular ion as a function of the frequency of a probe laser, which passes through the ion trap, intersecting with the ion cloud (Lemaire et al., 2002; Oh et al., 2005; Polfer, 2011). Thus, IRMPD spectroscopy is a type of action spectroscopy whereby molecular fragmentation is interpreted as a signature of photon absorption. A detailed description of the technique is available in references (Aleese et al., 2006) and (Macaleese and Maître, 2007). By probing in the IR region, one obtains information on the frequencies of fundamental vibrational transitions, which may then be compared with the harmonic (and sometimes anharmonically-corrected) vibrational frequency predictions of electronic structure software packages. This, in turn, facilitates structural assignment based on the similarity between computed and measured spectra, and the identification of distinguishing/diagnostic spectral features.

Spectroscopic investigation of amino acids and amino acid-containing clusters continues to be an active field of research owing to the biological relevance of these systems (Nanita and Cooks, 2006; Mino et al., 2011; Stedwell et al., 2013; Sunahori et al., 2013; Armentrout et al., 2014; Seo et al., 2017, 2018; Heiles et al., 2018; Jašíková and Roithová, 2018; Ma et al., 2018; Scutelnic et al., 2018). In particular, serine has received a great deal of attention owing to the implication of the serine octamer in homochiral genesis (i.e., the origin of L-amino acid chiral preference in nature) (Counterman and Clemmer, 2001; Sunahori et al., 2013; Seo et al., 2017; Scutelnic et al., 2018). Indeed, the Bowers and von Helden groups recently published a series of high-profile studies detailing the assignment of the IR spectra for cryogenically-cooled protonated serine octamer, $[\mathrm{Ser}_8 + \mathrm{H}]^+$, and protonated serine dimer, $[\mathrm{Ser}_2 + \mathrm{H}]^+$ (Seo et al., 2017, 2018; Scutelnic et al., 2018). To demonstrate the utility of our augmented BH approach for searching PESs and assigning IR spectra, we employed our methodology to study $[\mathrm{Ser}_2 + \mathrm{H}]^+$.

To begin, preliminary B3LYP/6-311++G(d,p) optimizations were conducted for neutral and protonated serine monomers to obtain partial charges for utilization with the molecular mechanics force field. For neutral monomers, both canonical and zwitterionic initial guesses were employed, and only the canonical structures were obtained. For the protonated isomers, initial guesses protonated at the carbonyl group, the amine group, and the side-chain hydroxyl group were optimized; all resulted in an amine-protonated structure, in agreement with previously published results (Noguera et al., 2001). After the optimizations, the atomic partial charges were calculated using the CHelpG partition scheme to reproduce the electrostatic potential at the near exterior of the van der Waals radial surface (Breneman and Wiberg, 1990). DFT optimizations were run in parallel, threaded across 8 cores, and required approximately 1 hour per calculation. Following pre-optimization and partial charge calculations for the monomers, both moieties were combined to produce the protonated dimer for treatment with the BH code. To search the potential energy landscape, dihedral angles in both moieties were given random rotations of $-5^\circ \leq \phi \leq +5^\circ$ on each iteration of the BH algorithm. The neutral moiety was also given random rotations of $-5^\circ \leq \theta \leq +5^\circ$ around its body-fixed *x*–, *y*–, and *z* –axes, and random translation of -0.5 Å $\leq \eta \leq +0.5$ Å in each of the *x*–, *y*–, and *z*–directions. This ensures that the relative orientations of the two moieties are also sampled. For geometry optimization, the custom-written BH code interfaces with the Gaussian software package where the AMBER force-field is used as the model potential (Wang et al., 2006; Frisch et al., 2009). Following an initial run of 1,000 steps at a thermal energy of E ≈ 0.43 eV (T = 5,000 K) to generate candidate structures, several parallel BH runs of 10,000 steps were run at a thermal energy of E ≈ 0.09 eV (T = 1,000 K) to search the PES. In total, more than 60,000 cluster geometries were sampled.

To benchmark the augmented BH algorithm, eight standard BH simulations of 5,000 steps were conducted and structural

TABLE 2 | The results of eight (BH + interpolation) simulations of $[Ser_2 + H]^+$.

Simulation	# Isomers found		Global minimum (Hartree)	
	BH	+ Interpolation	BH	Interpolation
1	70	16	**−0.25984**	−0.25940
2	74	19	**−0.25984**	−0.25593
3	60	8	**−0.25984**	−0.25572
4	67	22	−0.25969	**−0.25984**
5	62	32	−0.25969	**−0.25984**
6	76	17	**−0.25984**	−0.25967
7	67	6	**−0.25984**	−0.25515
8	51	28	−0.25969	−0.25809

Geometries were optimized at the PM7 level of theory (Frisch et al., 2016). Bold values emphasize the lowest energy value among all the isomers obtained from the total of 8 searches.

interpolation was subsequently applied to the unique isomers identified at the PM7 level of theory. Unique $[Ser_2 + H]^+$ isomers were identified based on energetic differences ($\Delta E \geq 10^{-5}$ Hartree) and by using a value of 50.0 Å as the similarity threshold between isomer pairs within the Euclidean distance matrix (*vide supra*). Isomer pairs with Euclidean distances of more than 150.0 Å were candidates for structural interpolation. Due to the large number of potential isomer pairs (∼6,000 for each BH simulation), we chose to randomly select only 300 pairs to test the interpolation methodology. For each pair, the midpoint structure ($\lambda = 0.5$) was located as described above and optimized at the PM7 level of theory. The optimized geometry of the interpolated structure was then compared to those in the original BH set using the same energy and Euclidean distance thresholds as employed previously. The results of the eight parallel (BH + interpolation) simulations are summarized in **Table 2**.

There are two observations worth noting in **Table 2**. Firstly, the isomer sets that were identified by the standard BH algorithm are augmented considerably by post-simulation interpolation; on average 19 new isomers were identified by interpolating between the 300 randomly selected isomer pairs found by standard BH simulations. Secondly, although the global minimum structure was identified in only five of the eight standard BH simulations of 5,000 steps, introducing post-simulation interpolation improved the rate of identifying the $[Ser_2 + H]^+$ global minimum to seven out of eight simulations.

Following BH simulation, the 200 unique lowest energy structures were carried forward to re-optimization at the B3LYP/6-311++G(d,p) + GD3 level of theory (Becke, 1988, 1993; Grimme et al., 2010). This treatment reduced the total number of unique isomers to 40. To ensure that these structures were local minima on the PES (i.e., no negative eigenvalues in the Hessian matrix, rather than TSs which have one negative Hessian eigenvalue), harmonic frequency calculations were undertaken. These calculations also served to predict the vibrational (*viz.* IR) spectra of the isomers and to estimate thermochemical corrections (see sections 1.1 and 1.2 of the **Supplementary Materials** for details). Using the optimized geometries from the density functional theory calculations, the distance matrix (as described in Equations 2, 3) was constructed. Linkages for hierarchical clustering were then determined using Ward's minimum variance method as implemented in the Orange software package (https://orange.biolab.si/) (Demsar et al., 2013), which at each step finds the pair of clusters that leads to the minimum increase in total within-cluster variance after merging (Ward, 1963). The resulting dendrogram, which is plotted in **Figure 4**, clearly shows four distinct groups of geometric structures; these groups are highlighted in blue, red, green, and orange. To better visualize the data, we have also used multi-dimensional scaling to create a 2D plot of the clustered data (Wickelmaier, 2003; Borg and Groenen, 2005). Based on this hierarchical clustering analysis, we clearly see that the BH algorithm identified several local minima associated with four distinct regions of the $[Ser_2 + H]^+$ PES. The lowest energy isomer in each of these four regions (*viz.* isomers 1, 6, 14, and 22) are highlighted and labeled on the MDS plot. This type of analysis provides insight with respect to how thoroughly a region of the PES has been searched. For example, if only one or two data points were identified in the blue region of the MDS plot, one might decide to initialize an additional BH run starting from one of the previously identified geometries. Moreover, this analysis can help guide interpolation efforts to identify TSs or geometries associated with stable intermediates between two previously identified minima. For example, upon inspection of the MDS plot shown in **Figure 4**, one can identify two outliers associated with the red group (in the top left of the red section) and one outlier associated with the green group (bottom left of the green section). In principle, one might choose to explore the region between these features and the more closely clustered structures on the MDS plot via the methods described in section Interpolating Intermediate Geometries. We choose not to do so here, however, because these three structures are associated with isomers 38, 39, and 40 (the highest energy species in our set).

Having identified four low energy geometric groupings associated with the $[Ser_2 + H]^+$ PES, we can then visually inspect the structures to rationalize their association via hierarchical clustering. In doing so, we find that the clustered species are associated with four distinct binding motifs, which we label motifs 1 (orange), 2 (blue), 3 (green), and 4 (red). The 3D structures and 2D chemical structures for the lowest energy isomer in each group is provided in **Figure 5**. Motifs 1 and 3 are associated with bidentate complexation between the ammonium group of the protonated moiety and the neutral moiety. In the case of motif 1, the ammonium group forms intermolecular hydrogen bonds with the amino group and the hydroxyl group of the neutral moiety. In contrast, motif 3 forms intermolecular hydrogen bonds with the hydroxyl group and the carboxylic acid group of the neutral moiety. Motifs 2 and 4 are associated with monodentate complexation between the ammonium group of the protonated moiety and the neutral moiety. These two binding motifs differ in terms of the relative orientations of the two serine moieties and with respect to the presence of a O–H•••N intramolecular hydrogen bond (IMHB) in the neutral moiety (motif 2) versus a O–H•••O IMHB in the neutral moiety (motif 4).

To determine which (if any) of the computed $[Ser_2 + H]^+$ isomers are observed experimentally, calculated harmonic vibrational spectra were compared against the experimental

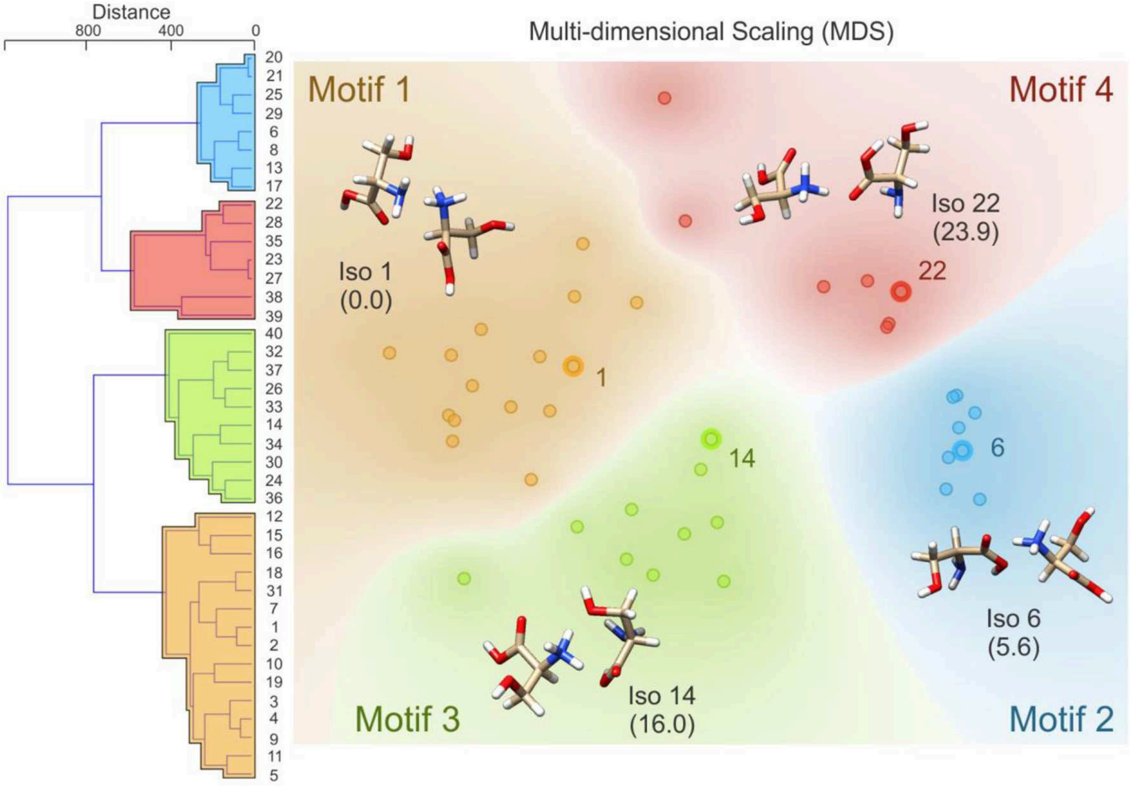


FIGURE 4 | (Left) The distance dendrogram for the protonated serine dimer. Isomer numbers are indicated for each branch of the dendrogram. **(Right)** A multi-dimensional scaling 2D projection of the hierarchical clustered data. Isomers are numbered in order of increasing energy above the global minimum (isomer 1). Standard Gibbs energies (in parentheses) are reported in kJ mol^{-1}. Calculations were conducted at the B3LYP/6-311++G(d,p) + GD3 level of theory as implemented in Gaussian 09 (Frisch et al., 2009).

IRMPD spectrum using the methodology outlined by Fu and Hopkins (2018) The experimental spectrum employed was a concatenation of the spectra recorded by Seo et al. in the 1,000–1,900 cm^{-1} region and by Sunahori et al. in the 3,200–3,800 cm^{-1} region (Sunahori et al., 2013; Seo et al., 2018). These spectra were digitized using a custom-written python script from figures in their respective publications, interpolated in 2 cm^{-1} intervals, then normalized such that the maximum intensity in each region was set to 1. Calculated IR spectra were first scaled using appropriate frequency scaling factors and broadened with a Lorentzian line shape of 15 cm^{-1} FWHM (Andersson and Uvdal, 2005; Fu and Hopkins, 2018), and then were similarly interpolated and normalized. The intensity vectors (i.e., y-values) of the computed spectra were then compared with the experimental spectrum by taking the Euclidian distance (d_{Euc}) between the intensity vectors and assigning a scaled similarity index as per:

$$Scaled\ Similarity = 1 - \frac{\left(d_{Euc} - d_{Euc}^{Min}\right)}{\left(d_{Euc}^{Max} - d_{Euc}^{Min}\right)} \qquad (11)$$

Where d_{Euc}^{Min} is the minimum Euclidean distance amongst the set of vectors and d_{Euc}^{Max} is the maximum Euclidean distance amongst the set of vectors following subtraction of the minimum distance. This treatment generates a scaled similarity index that ranges between 0 (worst match) and 1 (best match). The scaled similarities for the computed $[Ser_2 + H]^+$ isomer spectra are plotted in **Figure 6**. Inspection of **Figure 6** indicates that Isomer 6 yields a significantly better match to the experimental spectrum than do other isomers. Moreover, we find that four of the five best matches are provided by isomers associated with binding motif 2. This suggests that, despite the fact that motif 1 is associated with the lowest energy region of the $[Ser_2 + H]^+$ PES at $T = 298\,K$ and $P = 1$ atm, the region of the PES associated with motif 2 is predominantly populated in ion trap experiments.

Figure 7 plots the experimental IRMPD spectrum for $[Ser_2 + H]^+$ and the computed spectra for isomers 1, 6 (best match), 14, and 22—the lowest energy isomers associated with each of the four binding motifs. The diagnostic peaks, which are highlighted in blue in **Figure 7**, are associated with the HNH angle bending motions (*ca.* 1,450 cm^{-1}) and N–H bond stretching motions (*ca.* 3,250 cm^{-1}) of the ammonium and amino groups. Although isomer 1 is the global minimum structure based on standard Gibbs energies, the spectrum of isomer 6 (+5.6 kJ mol^{-1}) is much more representative of the experimental spectrum. This was also noted by Sunahori et al., who identified isomer 6 in their study (Sunahori et al., 2013). Kong et al. also identified isomer 6 in

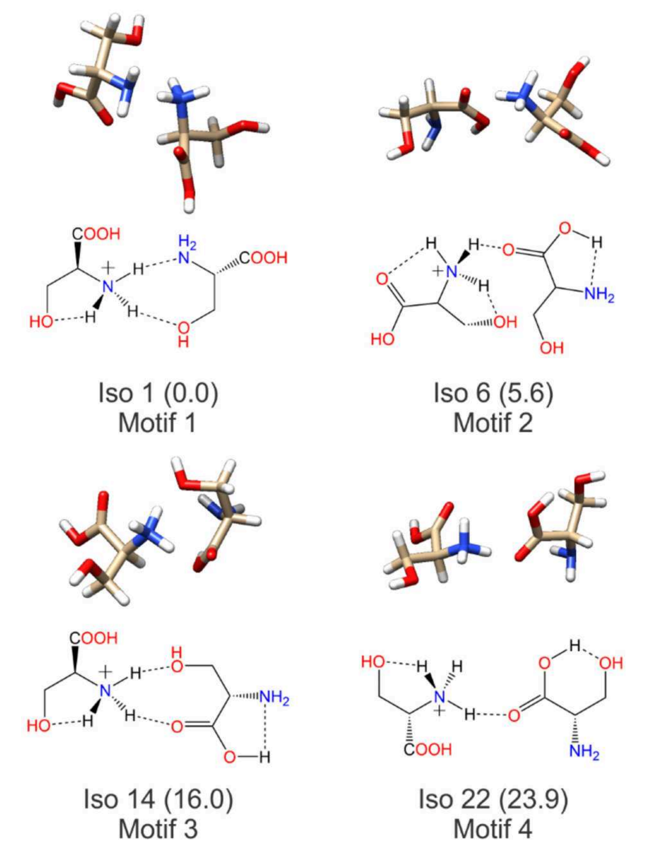


FIGURE 5 | The lowest energy isomers for each low energy binding motif of the protonated serine dimer. Motifs 1 and 3 show bidentate coordination between the two moieties, whereas motifs 2 and 4 exhibit monodentate coordination between the two moieties.

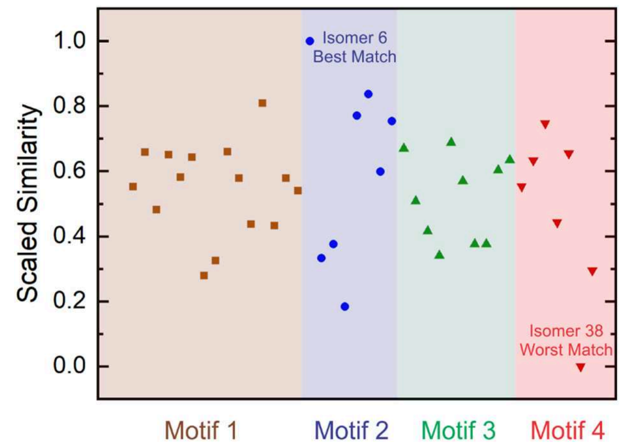


FIGURE 6 | Scaled Euclidean similarities of computed harmonic vibrational spectra to experimental IRMPD spectra for the protonated serine dimer. Isomer 6 gives the best match and Isomer 38 gives the worst match amongst the 40-isomer set. Isomers are ordered in increasing energy from left to right in each motif.

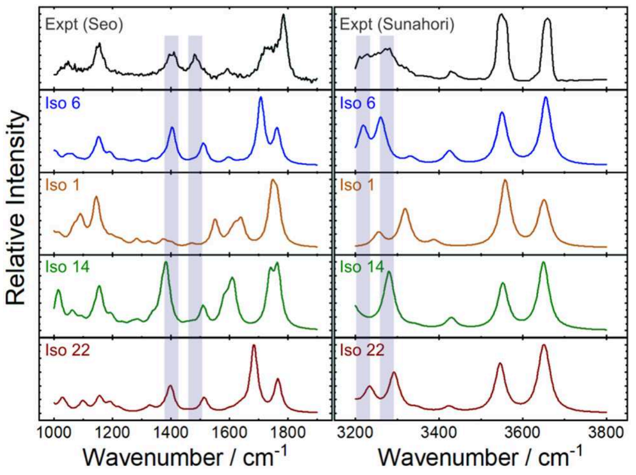


FIGURE 7 | Experimental IRMPD spectra and computed harmonic vibrational spectra for the protonated serine dimer. The experimental spectra were adapted from Seo et al. (2018) and Sunahori et al. (2013). The computed IR spectra are associated with the lowest energy isomer for each of the four binding motifs. Scaling factors of 0.9679 and 0.95 were employed for the 1,000–1,900 cm^{-1} and 3,200–3,800 cm^{-1} regions, respectively (Andersson and Uvdal, 2005; Fu and Hopkins, 2018).

their work, but apparently did not consider it in their spectral assignment (Kong et al., 2006). Note that harmonic spectra were scaled by 0.9679 in the 1,000–2,000 cm^{-1} region and 0.95 in the 3,000–4,000 cm^{-1} region, as recommended by NIST and based on previous work for similar systems (Andersson and Uvdal, 2005; Fu and Hopkins, 2018).

It is necessary to highlight three caveats for the above example of identifying the spectral carrier of $[Ser_2 + H]^+$. First, to create the experimental spectrum that we used in our assignment, we collated the results of two separate studies (Sunahori et al., 2013; Seo et al., 2018). It is not necessarily true that the same ensemble populations were produced under the experimental conditions employed in both of these studies. However, given that isomer 6 provides the best match to both regions of the experimental spectrum, it seems to be that instrument conditions were similar in these two cases. A second consideration is the fact that peak intensities in IRMPD spectra are not necessarily well-modeled by computed absorption spectra owing to the fact that IRMPD intensities are dependent on absorption cross sections *and* the coupling efficiency for accessing dissociative channels. (Parneix et al., 2013) The methodology outline above assumes that the computed linear absorption intensities are representative of IRMPD intensities or, barring that, that the IRMPD intensities for a given band vary similarly from the computed intensity for all isomeric species. Finally, the above treatment also assumes that the computed harmonic frequencies suitably model the experimental spectrum. The validity of this assumption depends on the accuracy of the model chemistry and on the anharmonicity of the system being studied. While the $[Ser_2 + H]^+$ is apparently well-modeled by the B3LYP/6-311++G(d,p) + GD3 approach employed here, one should in general be aware of the anharmonic nature of hydrogen bonds and shared protons (Schofield et al., 2005; Oomens et al., 2009; Steill et al., 2011; Ieritano et al., 2016).

Case Study 2: Dynamic Collision Cross Section of Protonated Alanine Tripeptide

Ion mobility spectrometry (IMS) is widely employed in the detection of illicit substances and for structural elucidation of ions (Collins and Lee, 2002; Verkouteren and Staymates, 2011; Lapthorn et al., 2013; Lanucara et al., 2014; Cumeras et al., 2015; Cajka and Fiehn, 2016; Paglia and Astarita, 2017). The success of IMS in determining analyte structure relies on accurate modeling of ion structure and subsequent calculation of CCSs for comparison with those determined experimentally. Experimental CCSs are obtained by relating the ion mobility, K, to CCS via the Mason-Schamp Equation (Mason and Mcdaniel, 1988; Ieritano et al., 2019b):

$$K = \frac{\sqrt{18\pi}}{16}\sqrt{\frac{1}{m_{ion}} + \frac{1}{m_{gas}}}\frac{ze}{\sqrt{k_b T}}\frac{1}{\Omega_{avg}}\frac{1}{N} \tag{12}$$

Where m_{gas} is the mass of the buffer gas, N is the number density of the gas, m_{ion} is the mass of the ion, z is the ion charge state, e is the elementary charge, k_b is the Boltzmann constant, T is the temperature, and Ω_{avg} is the orientationally-averaged CCS. Typically, ion structures are viewed as rigid and ensembles are approximated as being composed of only a single structure in cases where multiple distinct signals are unresolved. This view is somewhat tenuous, particularly in the differential mobility spectrometry (DMS) variant of IMS wherein rapidly oscillating electric field conditions drive separations based on mobility differences between the high- and low-field portions of the applied waveform (Guevremont and Purves, 1999; Guevremont, 2004; Krylov et al., 2007, 2009; Krylov and Nazarov, 2009; Hopkins, 2015, 2019). The phenomenon of differential ion mobility is still not well-understood, and there is as yet no first principles model (Guevremont and Purves, 1999; Guevremont, 2004; Krylov et al., 2007, 2009; Krylov and Nazarov, 2009; Hopkins, 2015, 2019). However, one can view the effective temperature of an analyte ion in terms of the changing field conditions; the ion is relatively cold under low-field conditions and relatively hot under high-field conditions (Viehland and Mason, 1995; Robinson et al., 2008; Hopkins, 2019). By estimating ion temperatures with two-temperature theory (Robinson et al., 2008; Siems et al., 2016), we find that field-induced heating leads to effective ion temperature variations in the range of 300–800 K during one duty cycle of the commonly applied maximum electric field in the DMS cell (Hopkins, 2019). The variation in electric field, and therefore effective ion temperature, affects the ion mobility in two ways, the most obvious being the reduction of mobility with increasing temperature as predicted by Equation (12). Somewhat more subtle is the fact that Ω_{avg} must also be temperature-dependent since at elevated temperatures ions are able to access a larger region of the associated PES (assuming equipartition amongst the various DoFs of the molecule). Consequently, to accurately model an ion's Ω_{avg}, one must identify which geometric structures are accessible under the given experimental conditions and estimate the contribution of that structure to the time-averaged CCS of the ion.

If we consider the case of protonated alanine tripeptide, $[\mathrm{AAA} + \mathrm{H}]^+$, there are several internal DoFs associated with dihedral angle rotations that can yield a variety of conformations. Upon application of the BH algorithm to search the PES of the $[\mathrm{AAA} + \mathrm{H}]^+$ molecular ion, followed by re-optimization of the candidate structures at the B3LYP/6-31++G(d,p) + GD3 level of theory (Becke, 1988, 1993; Grimme et al., 2010), fourteen low energy conformations were identified. These structures are shown in **Figure 8** along with their relative standard Gibbs energies (in kJ mol^{-1}) (see sections 2.1 and 2.2 of the **Supplementary Materials** for details). Calculating the cosine distances between the various mass-weighted distance vectors and subsequent application of WPGMA agglomerative hierarchical clustering yields the dendrogram plot shown in **Figure 8**. Five unique sets of conformers are highlighted in the dendrogram. The set highlighted in yellow, of which the global minimum conformer is a member, contains compact structures that are stabilized by an IMHB between the protonated N-terminus and the carbonyl oxygen atom of the C-terminus. The set highlighted in green also contains relatively compact structures, but hydrogen bonding instead occurs between the protonated N-terminus and the hydroxyl group of the C-terminus. The set highlighted in red, on the other hand, contains elongated structures (i.e., the N- and C-termini do not interact). Conformers 6 and 9 (blue and orange, respectively) are intermediate species between the compact species (yellow and green sets) and the elongated species (red set). In the case of conformer 6, the N-terminus forms an IMHB with the nearest amide carbonyl rather than with the C-terminus. In contrast, the C-terminus of conformer 9 forms an IMHB with the most distant amino nitrogen instead of with the N-terminus.

If we calculate the relative Gibbs energies of the $[\mathrm{AAA} + \mathrm{H}]^+$ conformers as a function of temperature, an interesting picture emerges. Owing to differences in the entropic contributions to the Gibbs energies, at low temperature the compact, H-bonded conformers associated with the yellow group are the dominant species in the ensemble, whereas at high temperature the elongated, non-H-bonded species in the red group dominate. One can estimate the relative populations of the various conformers via (Oh and Zeng, 1999; Vehkamäki, 2006; Hopkins, 2019):

$$N_i = N_0 e^{-\frac{\Delta G_{rel}}{k_B T}} \tag{13}$$

Where N_0 is the relative population of the lowest energy cluster (usually set to 1), N_i is the relative population of the ith cluster, ΔG_{rel} is the Gibbs energy of formation relative to the lowest energy cluster, and k_B is Boltzmann's constant. By calculating the relative populations of the clusters as a function of temperature (at a constant pressure of P = 1 atm), one can produce a temperature-dependent relative population plot as shown in **Figure 9**.

Figure 9 shows that at *ca*. T = 420 K $[\mathrm{AAA} + \mathrm{H}]^+$ conformer 3 becomes the most populated species in the ensemble (i.e., the global minimum structure on the Gibbs energy surface). As

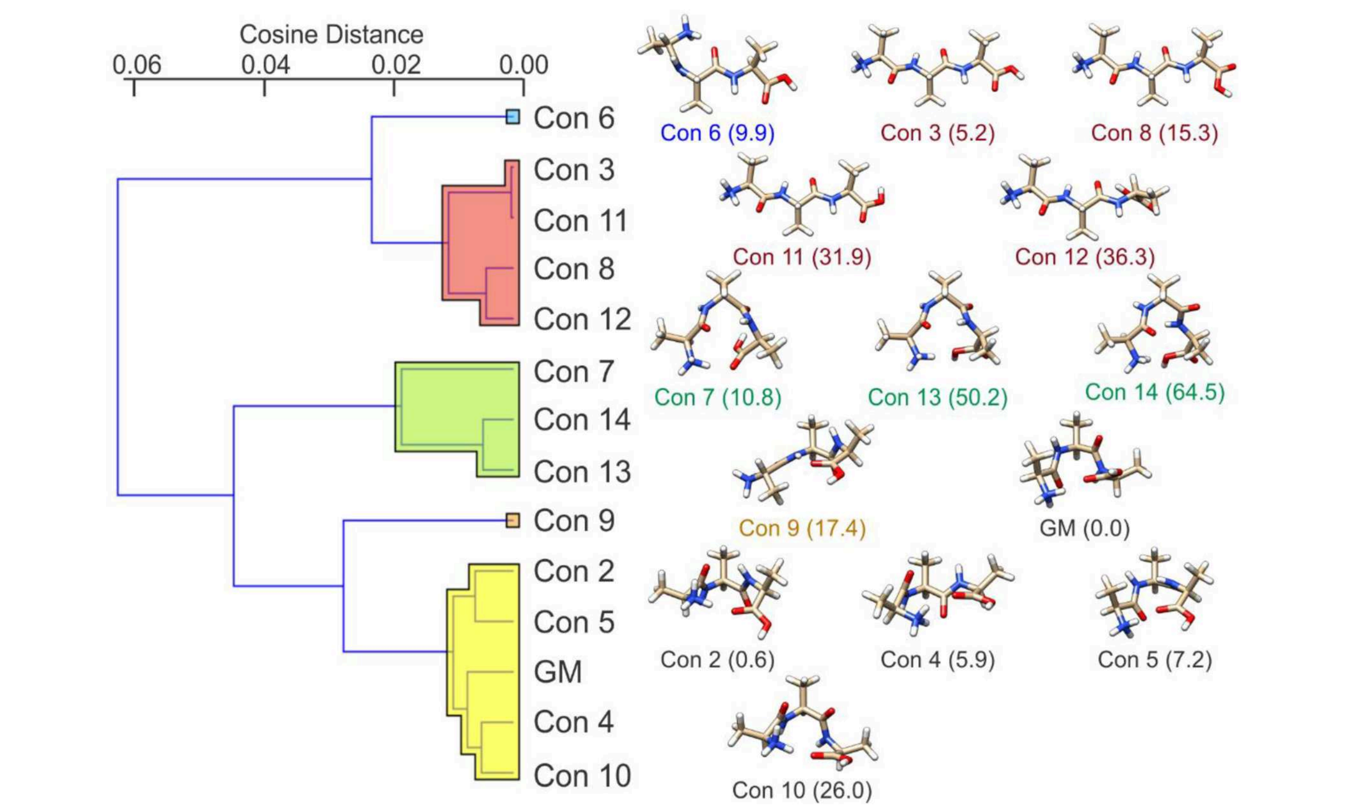


FIGURE 8 | (Left) The cosine distance dendrogram for protonated alanine tripeptide, $[AAA+H]^+$. **(Right)** Molecular geometries and relative standard Gibbs energies (kJ mol^{-1}; in parentheses). Calculations were conducted at the B3LYP/6-311++G(d,p) + GD3 level of theory as implemented in Gaussian 16 (Frisch et al., 2016). Conformers are numbered in order of increasing energy relative to that of the global minimum (GM) structure.

the temperature increases further, conformer 3 is increasingly stabilized with respect to conformers 1 (the low T global minimum) and 2. At temperatures above 660 K, conformers 1 and 2 become minor contributors to the overall ensemble population in favor of conformers 3 and 8. This "tilting" of the Gibbs energy landscape as a function of temperature essentially decants the conformers associated with the yellow set into the red set (see **Figure 8**) as field-induced ion temperature increases, and back again as the temperature decreases during the low field portion of the oscillating DMS waveform. This dynamic process of peptide unfolding and re-folding yields a dynamic temperature-dependent ion CCS that, along with the effect of increased carrier gas viscosity at higher temperature (Mason and Mcdaniel, 1988; Hopkins, 2019), gives rise to differential mobility behavior. If one assumes that the ion quickly reaches thermal equilibrium, which is likely given the conditions of the DMS cell (1 atm of carrier gas), one can estimate the temperature-dependent ion CCS as a sum of the Boltzmann-weighted conformer CCSs (Ieritano et al., 2019a). This is plotted for $[AAA + H]^+$ in **Figure 10**. It is worth noting that the experimentally-measured T ≈ 293 K value of $\Omega_{ave}(N_2) = 151$ Å^2 (Bush et al., 2012) is well-modeled by the T = 300 K Boltzmann-weighted sum of the various isomer CCSs as calculated using the MobCal-MPI code (https://uwaterloo.ca/hopkins-lab/mobcal-mpi), $\Omega_{Boltzmann}(N_2) = 151.3$ Å^2 (Ieritano et al., 2019b). In comparison, the calculated CCS for the static global minimum structure is $\Omega_{Boltzmann}(N_2) = 148.7$ Å^2. This demonstrates that even at a relatively low fixed temperature, there is some benefit in considering the relative populations of the conformeric species present in the experimental ensemble.

SUMMARY

Because the PESs of complex, fluxional molecular systems tend to be characterized by multiple funnels (*viz.* collections of closely related local minima), the BH framework has proven to be an effective search and optimization strategy (Locatelli, 2005; Olson et al., 2012). However, owing to the stochasticity of the algorithm, which is predominantly due to the random perturbative component, it is sometimes useful to introduce additional criteria which limit the regions of exploration on the PES. This has been traditionally accomplished by exploring specific degrees of freedom (e.g., dihedral rotations) on the potential energy landscape and by introducing a thermal energy distribution as a probabilistic means of accepting/rejecting random geometric perturbations. We have also introduced techniques from unsupervised machine learning, specifically distance matrices and hierarchical clustering, to further augment the BH algorithm. Although currently implemented as a separate module, these machine learning augmentations will in the

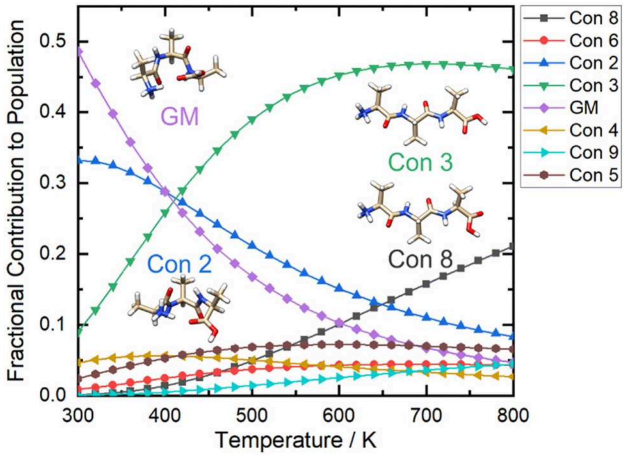


FIGURE 9 | The relative populations of the low energy conformers of protonated alanine tripeptide, $[AAA{+}H]^+$, as estimated via Gibbs energy calculations over the temperature range T = 300–800 K. Calculations were conducted at the B3LYP/6-311++G(d,p) + GD3 level of theory as implemented in Gaussian 16 (Frisch et al., 2016). Conformers are numbered in order of increasing energy relative to that of the global minimum (GM; i.e., conformer 1) structure.

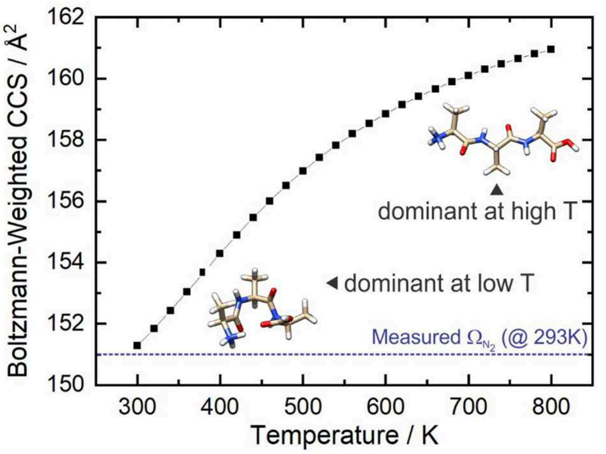


FIGURE 10 | The Boltzmann-weighted CCS of $[AAA + H]^+$ as a function of temperature at P = 1 atm. The dashed blue line shows the orientationally-averaged CCS, Ω_{ave}, measured in N_2 at room temperature (T ≈ 293 K) (Bush et al., 2012).

future be incorporated for on-the-fly geometric analyses, which would ultimately provide additional control and efficiency during execution of the search algorithm afforded by reducing the search space to pertinent regions connecting known stationary points. This is particularly useful in identifying intermediate local minima and TSs between known isomers. Moreover, utilizing these same methods post-BH provides deep insights into the relation between stationary points and how these are partitioned on the potential energy landscape. This can be of great benefit in modeling experimental ensembles and in rationalizing the observation of kinetically-trapped species and dynamic molecular geometries.

In this manuscript we highlight the power of the BH framework in two case studies: (1) assigning the spectral carrier(s) of the IRMPD spectrum of $[Ser_2{+}H]^+$ and (2) modeling the temperature-dependent collision cross sections of $[AAA{+}H]^+$. In case study 1, we show that a thorough mapping of the potential energy landscape is warranted to identify the species probed in gas phase ion spectroscopic studies of weakly-bound clusters. In the case of the protonated serine dimer, rather than observing the lowest energy isomer (as expected based on standard Gibbs energies), Seo et al. and Sunahori et al. observed a species that was associated with a relatively remote, higher energy region of the cluster PES (Sunahori et al., 2013; Seo et al., 2018). It is still an open question as to whether this was due to kinetic trapping during production or formation of this species *in situ* due to field-induced heating within the ion traps. In case study 2, we show that mapping PESs to identify low energy conformer geometries, which were subsequently refined at a higher level of quantum chemical theory, provides insight into how molecular geometry changes with increasing temperature. For $[AAA{+}H]^+$, increasing the temperature of the system results in the dissociation of IMHBs and the formation of larger elongated structures compared to the compact H-bonded species favored at low temperature. We also demonstrate that modeling molecular collision cross sections as a Boltzmann-weighted sum of the CCSs for accessible conformers provides an accurate estimate of those measured experimentally (0.3 Å^2 difference). It should be noted that this treatment assumes that the accessible conformers are readily interconvertible, and that thermal equilibrium is quickly established. In principle, one could also employ the interpolation techniques described in section Interpolating Intermediate Geometries to calculate barriers to interconversion and validate this assumption. However, the fact that our calculations yield results that are in excellent agreement with experimental measurements indicates that, in this case, the assumption is valid.

Ultimately, the BH framework is a useful approach to characterizing the structures and dynamics of chemical systems which exhibit PESs of high dimensionality. Examples of such systems range from weakly-bound nanoclusters to biological macromolecules. We note that, despite the success of our current implementation, the development of the BH framework by ourselves and others is ongoing. We expect that further tuning will improve general performance and, owing to the versatility of the method, that BH performance for specific tasks will continue to improve by tailoring key features of the algorithm.

AUTHOR CONTRIBUTIONS

CZ conducted the serine dimer work and wrote the original draft of manuscript. CI conducted the alanine tripeptide work. WH wrote the final draft of the manuscript.

FUNDING

WH acknowledges funding from the Natural Sciences and Engineering Research Council (NSERC) of Canada in the form of a Discovery Grant and from the Province of Ontario in the form of an Early Researcher Award (ERA). CI acknowledges funding from NSERC in the form of a post graduate scholarship.

ACKNOWLEDGMENTS

The authors would like to acknowledge Amy M. -C. Yang for help with data acquisition.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fchem.2019.00519/full#supplementary-material

Calculated energies and atomic XYZ coordinates for all species are available as **Supplementary Material**.

REFERENCES

Aleese, L. M., Simon, A., Mcmahon, T. B., Ortega, J.-M., Scuderi, D., Lemaire, J., et al. (2006). Mid-IR spectroscopy of protonated leucine methyl ester performed with an FTICR or a Paul type ion-trap. *Int. J. Mass Spectrometry* 249–250, 14–20. doi: 10.1016/j.ijms.2006.01.008

Andersson, M. P., and Uvdal, P. (2005). New scale factors for harmonic vibrational frequencies using the B3LYP density functional method with the triple-ζ basis set 6-311+G(d,p). *J. Phys. Chem. A* 109, 2937–2941. doi: 10.1021/jp045733a

Armentrout, P. B., Yang, B., and Rodgers, M. T. (2014). Metal cation dependence of interactions with amino acids: Bond dissociation energies of Rb+ and Cs+ to the acidic amino acids and their amide derivatives. *J. Phys. Chem. B* 118, 4300–4314. doi: 10.1021/jp5001754

Becke, A. D. (1988). Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys. Rev. A* 38, 3098–3100. doi: 10.1103/PhysRevA.38.3098

Becke, A. D. (1993). Density-functional thermochemistry.3. the role of exact exchange. *J. Chem. Phys.* 98, 5648–5652. doi: 10.1063/1.464913

Borg, I., and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling.* New York, NY: Springer.

Breneman, C. M., and Wiberg, K. B. (1990). Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* 11, 361–373. doi: 10.1002/jcc.540110311

Bush, M. F., Campuzano, I. D. G., and Robinson, C. V. (2012). Ion mobility mass spectrometry of peptide ions: effects of drift gas and calibration strategies. *Analyt. Chem.* 84, 7124–7130. doi: 10.1021/ac3014498

Cajka, T., and Fiehn, O. (2016). Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. *Analyt. Chem.* 88, 524–545. doi: 10.1021/acs.analchem.5b04491

Call, S. T., Zubarev, D. Y., and Boldyrev, A. I. (2007). Global minimum structure searches via particle swarm optimization. *J. Comput. Chem.* 28, 1177–1186. doi: 10.1002/jcc.20621

Campbell, J. L., Le Blanc, J. C. Y., and Schneider, B. B. (2012). Probing electrospray ionization dynamics using differential mobility spectrometry: the curious case of 4-aminobenzoic acid. *Analyt. Chem.* 84, 7857–7864. doi: 10.1021/ac301529w

Campbell, J. L., Yang, A. M.-C., Melo, L. R., and Hopkins, W. S. (2016). Studying gas-phase interconversion of tautomers using differential mobility spectrometry. *J. Am. Soc. Mass Spectrom.* 27, 1277–1284. doi: 10.1007/s13361-016-1392-2

Cereto, M. (2015). Molecular fingerprint similarity search in virtual screening. *Methods* 71, 58–63. doi: 10.1016/j.ymeth.2014.08.005

Choudhary, B. (2003). *The Elements of Complex Analysis, 2nd Edn*, ed K. K. Gupta (New Delhi: New Age International Limited Publishers).

Collins, D., and Lee, M. (2002). Developments in ion mobility spectrometry–mass spectrometry. *Analyt. Bioanalyt. Chem.* 372, 66–73. doi: 10.1007/s00216-001-1195-5

Counterman, A. E., and Clemmer, D. E. (2001). Magic number clusters of serine in the gas phase. *J. Phys. Chem. B* 105, 8092–8096. doi: 10.1021/jp011421l

Cumeras, R., Figueras, E., Davis, C. E., Baumbach, J. I., and Gràcia, I. (2015). Review on ion mobility spectrometry. Part 1: current instrumentation. *Analyst* 140, 1376–1390. doi: 10.1039/C4AN01100G

Day, W. H. E., and Edelsbrunner, H. (1984). Efficient algorithms for agglomerative hierarchical clustering methods. *J. Classification* 1, 7–24. doi: 10.1007/BF01890115

Demsar, J., Curk, T., Erjavex, A., Gorup, C., Hocevar, T., Milutinovic, M., et al. (2013). Orange: data mining toolbox in python. *J. Mach. Learn. Res.* 14, 2349–2353.

Eberhart, R., and Yuhui, S. (2001). "Particle swarm optimization: developments, applications and resources," in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546),* 81, 81–86.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09 Revision D.01.* Wallingford, CT: Gaussian, Inc.

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2016). *Gaussian 16 Rev. B.01.* Wallingford, CT.

Fritzke, B. (1994). "NIPS'94," in *Proceedings of the 7th International Conference on Neural Information Processing Systems*. Cambridge, MA: MIT Press, 625–632.

Fu, W., and Hopkins, W. S. (2018). Applying machine learning to vibrational spectroscopy. *J. Phys. Chem. A* 122, 167–171. doi: 10.1021/acs.jpca.7b10303

Gentle, J. E. (2007). *Matrix Algebra Theory, Computations, and Applications in Statistics.* New York, NY: Springer, 261–319.

Grimme, S., Antony, J., Ehrlich, S., and Krieg, H. (2010). A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* 132:154104. doi: 10.1063/1.3382344

Guevremont, R. (2004). High-field asymmetric waveform ion mobility spectrometry: a new tool for mass spectrometry. *J. Chromatogr. A* 1058, 3–19. doi: 10.1016/S0021-9673(04)01478-5

Guevremont, R., and Purves, R. W. (1999). Atmospheric pressure ion focusing in a high-field asymmetric waveform ion mobility spectrometer. *Rev. Sci. Instruments* 70, 1370–1383. doi: 10.1063/1.1149599

Heiles, S., Berden, G., Oomens, J., and Williams, E. R. (2018). Competition between salt bridge and non-zwitterionic structures in deprotonated amino acid dimers. *Phys. Chem. Chem. Phys.* 20, 15641–15652. doi: 10.1039/C8CP01458B

Heller, S., Mcnaught, A., Stein, S., Tchekhovskoi, D., and Pletnev, I. (2013). InChI - The worldwide chemical structure identifier standard. *J. Cheminform.* 5, 1–9. doi: 10.1186/1758-2946-5-7

Hopkins, W. S. (2015). Determining the properties of gas-phase clusters. *Mol. Phys.* 113, 3151–3158. doi: 10.1080/00268976.2015.1053545

Hopkins, W. S. (2019). "Chapter four - dynamic clustering and ion microsolvation," in *Comprehensive Analytical Chemistry, Vol. 83*, eds. W.A. Donald and J.S. Prell (Amsterdam: Elsevier), 83–122.

Hopkins, W. S., Marta, R. A., and Mcmahon, T. B. (2013). Proton-bound 3-cyanophenylalanine trimethylamine clusters: isomer-specific fragmentation pathways and evidence of gas-phase zwitterions. *J. Phys. Chem. A* 117, 10714–10718. doi: 10.1021/jp407766j

Hopkins, W. S., Marta, R. A., Steinmetz, V., and Mcmahon, T. B. (2015). Mode-specific fragmentation of amino acid-containing clusters. *Phys. Chem. Chem. Phys.* 17, 28548–28555. doi: 10.1039/C5CP03517A

Ieritano, C., Campbell, J. L., and Hopkins, W. S. (2019a). Unravelling the factors that drive separation in differential mobility spectrometry: a case study of regioisomeric phosphatidylcholine adduct. *Int. J. Mass Spectrom.* 444:116182. doi: 10.1016/j.ijms.2019.116182

Ieritano, C., Carr, P. J. J., Hasan, M., Burt, M., Marta, R. A., Steinmetz, V., et al. (2016). The structures and properties of proton- and alkali-bound cysteine dimers. *Phys. Chem. Chem. Phys.* 18, 4704–4710. doi: 10.1039/C5CP07414B

Ieritano, C., Crouse, J., Campbell, J. L., and Hopkins, W. S. (2019b). A parallelized molecular collision cross section package with optimized accuracy and efficiency. *Analyst* 144, 1660–1670. doi: 10.1039/C8AN02150C

Jašíková, L., and Roithová, J. (2018). Infrared multiphoton dissociation spectroscopy with free-electron lasers: on the road from small molecules to biomolecules. *Chem. A Eur. J.* 24, 3374–3390. doi: 10.1002/chem.201705692

Jones, D. R., Perttunen, C. D., and Stuckman, B. E. (1993). Lipschitzian optimization without the Lipschitz constant. *J. Optimization Theory Appl.* 79, 157–181. doi: 10.1007/BF00941892

Kabsch, W. (1976). A solution for the best rotation to relate two sets of vectors. *Acta Crystallograph. Section A* 32, 922–923. doi: 10.1107/S0567739476001873

Kennedy, J., and Eberhart, R. (1995). Particle swarm optimization. 4, 1942–1948.

Kim, Y., Choi, S., and Kim, W. Y. (2014). Efficient basin-hopping sampling of reaction intermediates through molecular fragmentation and graph theory. *J. Chem. Theory Comput.* 10, 2419–2426. doi: 10.1021/ct500136x

Kohonen, T. (1990). The self-organizing map. *Proc. IEEE* 78, 1464–1480. doi: 10.1109/5.58325

Kong, X., Tsai, I.-A., Sabu, S., Han, C.-C., Lee, Y. T., Chang, H.-C., et al. (2006). Progressive stabilization of zwitterionic structures in [H(Ser)2–8]+ studied by infrared photodissociation spectroscopy. *Angewandte Chemie Int. Ed.* 45, 4130–4134. doi: 10.1002/anie.200600597

Krylov, E. V., Coy, S. L., and Nazarov, E. G. (2009). Temperature effects in differential mobility spectrometry. *Int. J. Mass Spectrometry* 279, 119–125. doi: 10.1016/j.ijms.2008.10.025

Krylov, E. V., and Nazarov, E. G. (2009). Electric field dependence of the ion mobility. *Int. J. Mass Spectrometry* 285, 149–156. doi: 10.1016/j.ijms.2009.05.009

Krylov, E. V., Nazarov, E. G., and Miller, R. A. (2007). Differential mobility spectrometer: model of operation. *Int. J. Mass Spectrom.* 266, 76–85. doi: 10.1016/j.ijms.2007.07.003

Kumar, A., and Zhang, K. Y. J. (2018). Advances in the development of shape similarity methods and their application in drug discovery. *Front. Chem.* 6, 1–21. doi: 10.3389/fchem.2018.00315

Lanucara, F., Holman, S. W., Gray, C. J., and Eyers, C. E. (2014). The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* 6:281. doi: 10.1038/nchem.1889

Lapthorn, C., Pullen, F., and Chowdhry, B. Z. (2013). Ion mobility spectrometry-mass spectrometry (IMS-MS) of small molecules: separating and assigning structures to ions. *Mass Spectrometry Rev.* 32, 43–71. doi: 10.1002/mas.21349

Leary, R. H. (2000). Global optimization on funneling landscapes. *J. Global Optimiz.* 18, 367–383. doi: 10.1023/A:1026500301312

Lecours, M. J., Chow, W. C. T., and Hopkins, W. S. (2014). Density functional theory study of RhnS0,+/- and Rh-n+1(0,+/-) (n=1-9). *J. Phys. Chem. A* 118, 4278–4287. doi: 10.1021/jp412457m

Lemaire, J., Boissel, P., Heninger, M., Mauclaire, G., Bellec, G., Mestdagh, H., et al. (2002). Gas phase infrared spectroscopy of selectively prepared ions. *Phys. Rev. Lett.* 89:273002. doi: 10.1103/PhysRevLett.89.273002

Locatelli, M. (2005). On the multilevel structure of global optimization problems. *Comput. Optimization Appl.* 30, 5–22. doi: 10.1007/s10589-005-4561-y

Locatelli, M., and Schoen, F. (2013). Global optimization: theory, algorithms, and applications. *Soc. Industr. Appl. Mathematics*. 30, 5–22. doi: 10.1137/1.9781611972672

Ma, L., Ren, J., Feng, R., Zhang, K., and Kong, X. (2018). Structural characterizations of protonated homodimers of amino acids: revealed by infrared multiple photon dissociation (IRMPD) spectroscopy and theoretical calculations. *Chin. Chem. Lett.* 29, 1333–1339. doi: 10.1016/j.cclet.2018.02.008

Macaleese, L., and Maître, P. (2007). Infrared spectroscopy of organometallic ions in the gas phase: from model to real world complexes. *Mass Spectrometry Rev.* 26, 583–605. doi: 10.1002/mas.20138

Maeda, S., Taketsugu, T., and Morokuma, K. (2014). Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. *J. Computational Chem.* 35, 166–173. doi: 10.1002/jcc.23481

Martinez, T., and Schulten, K. (1991). *Artificial Neural Networks, Vol. 1*, eds T. Kohonen, K. Kakisara, O. Simula, J. Kangas (Amsterdam: Elsevier B.V.), 397–402.

Mason, E. A., and Mcdaniel, E. W. (1988). *Transport Properties of Ions in Gases.* New York, NY: John Wiley and Sons.

Michener, C. D., and Sokal, R. R. (1957). A quantitative approach to a problem in classification. *Evolution* 11, 130–162. doi: 10.1111/j.1558-5646.1957.tb02884.x

Mino, W. K., Gulyuz, K., Wang, D., Stedwell, C. N., and Polfer, N. C. (2011). Gas-phase structure and dissociation chemistry of protonated tryptophan elucidated by infrared multiple-photon dissociation spectroscopy. *J. Phys. Chem. Lett.* 2, 299–304. doi: 10.1021/jz1017174

Montavon, G., Hansen, K., Fazli, S., Rupp, M., Biegler, F., Ziehe, A., et al. (2012). "Learning invariant representations of molecules for atomization energy prediction," in Advances in Neural Information Processing Systems 25, eds F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc.), 440–448.

Nanita, S. C., and Cooks, R. G. (2006). Serine octamers: Cluster formation, reactions, and implications for biomolecule homochirality. *Angew. Chem. Int. Ed.* 45, 554–569. doi: 10.1002/anie.200501328

Noguera, M., RodríGuez-Santiago, L., Sodupe, M., and Bertran, J. (2001). Protonation of glycine, serine and cysteine. Conformations, proton affinities and intrinsic basicities. *J. Mol. Struc. Theochem*. 537, 307–318. doi: 10.1016/S0166-1280(00)00686-2

Oh, H. B., Lin, C., Hwang, H. Y., Zhai, H., Breuker, K., Zabrouskov, V., et al. (2005). Infrared photodissociation spectroscopy of electrosprayed ions in a Fourier transform mass spectrometer. *J. Am. Chem. Soc.* 127, 4076–4083. doi: 10.1021/ja040136n

Oh, K. J., and Zeng, X. C. (1999). Formation free energy of clusters in vapor-liquid nucleation: a Monte Carlo simulation study. *J. Chem. Phys.* 110, 4471–4476. doi: 10.1063/1.478331

Olson, B., Hashmi, I., Molloy, K., and Shehu, A. (2012). Basin hopping as a general and versatile optimization framework for the characterization of biological macromolecules. *Adv. Artificial Intellig.* 2012:19. doi: 10.1155/2012/674832

Oomens, J., Steill, J. D., and Redlich, B. (2009). Gas-phase IR spectroscopy of deprotonated amino acids. *J. Am. Chem. Soc.* 131, 4310–4319. doi: 10.1021/ja807615v

Paglia, G., and Astarita, G. (2017). Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.* 12:797. doi: 10.1038/nprot.2017.013

Parneix, P., Basire, M., and Calvo, F. (2013). Accurate modeling of infrared multiple photon dissociation spectra: the dynamical role of anharmonicities. *J. Phys. Chem. A* 117, 3954–3959. doi: 10.1021/jp402459f

Peng, C., Ayala, P. Y., Schlegel, H. B., and Frisch, M. J. (1996). Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.* 17, 49–56. doi: 10.1002/(SICI)1096-987X(19960115)17:1<49::AID-JCC5>3.0.CO;2-0

Peng, C., and Bernhard Schlegel, H. (1993). Combining synchronous transit and quasi-newton methods to find transition states. *Israel J. Chem.* 33, 449–454. doi: 10.1002/ijch.199300051

Piela, L., Olszewski, K. A., and Pillardy, J. (1994). On the stability of conformers. *Theochem. J. Mol. Struc.* 114, 229–239. doi: 10.1016/0166-1280(94)80105-3

Polfer, N. C. (2011). Infrared multiple photon dissociation spectroscopy of trapped ions. *Chem. Soc. Rev.* 40, 2211–2221. doi: 10.1039/c0cs00171f

Rahman, S. A., Bashton, M., Holliday, G. L., Schrader, R., and Thornton, J. M. (2009). Small Molecule Subgraph Detector (SMSD) toolkit. *J. Cheminform.* 1, 1–13. doi: 10.1186/1758-2946-1-12

Robinson, E. W., Shvartsburg, A. A., Tang, K., and Smith, R. D. (2008). Control of ion distortion in field asymmetric waveform ion mobility spectrometry via variation of dispersion field and gas temperature. *Analyt. Chem.* 80, 7508–7515. doi: 10.1021/ac800655d

Röder, K., and Wales, D. J. (2018). Mutational basin-hopping: combined structure and sequence optimization for biomolecules. *J. Phys. Chem. Lett.* 9, 6169–6173. doi: 10.1021/acs.jpclett.8b02839

Scheraga, H. A. (1992). Some approaches to the multiple-minima problem in the calculation of polypeptide and protein structures. *Int. J. Quantum Chem.* 42, 1529–1536. doi: 10.1002/qua.560420526

Schmidt, J., Meyer, M. M., Spector, I., and Kass, S. R. (2011). Infrared multiphoton dissociation spectroscopy study of protonated p-aminobenzoic acid: does electrospray ionization afford the amino- or carboxy-protonated ion? *J. Phys. Chem. A* 115, 7625–7632. doi: 10.1021/jp203829z

Schofield, D. P., Kjaergaard, H. G., Matthews, J., and Sinha, A. (2005). The OH-stretching and OOH-bending overtone spectrum of HOONO. *J. Chem. Phys.* 123:134318. doi: 10.1063/1.2047574

Scutelnic, V., Perez, M. A. S., Marianski, M., Warnke, S., Gregor, A., Seo, J., et al. (2018). The structure of the protonated serine octamer. *J. Am. Chem. Soc.* 140, 7554–7560. doi: 10.1021/jacs.8b02118

Seo, J., Hoffmann, W., Malerz, S., Warnke, S., Bowers, M. T., Pagel, K., et al. (2018). Side-chain effects on the structures of protonated amino acid dimers: a gas-phase infrared spectroscopy study. *Int. J. Mass Spectrometry* 429, 115–120. doi: 10.1016/j.ijms.2017.06.011

Seo, J., Warnke, S., Pagel, K., and Bowers, M. T. A. (2017). Infrared spectrum and structure of the homochiral serine octamer-dichloride complex. *Nat. Chem.* 9, 1263–1268. doi: 10.1038/nchem.2821

Shi, L. T., Wang, Z. Q., Hu, C. E., Cheng, Y., Zhu, J., and Ji, G. F. (2019). Possible lower energy isomer of carbon clusters C n (n = 11, 12) via particle swarm optimization algorithm: Ab initio investigation. *Chem. Phys. Lett.* 721, 74–85. doi: 10.1016/j.cplett.2019.02.028

Siems, W. F., Viehland, L. A., and Hill, H. H. (2016). Correcting the fundamental ion mobility equation for field effects. *Analyst* 141, 6396–6407. doi: 10.1039/C6AN01353H

Sokal, R. R., and Michener, C. D. (1958). A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.* 38, 1409–1438.

Stedwell, C. N., Galindo, J. F., Gulyuz, K., Roitberg, A. E., and Polfer, N. C. (2013). Crown complexation of protonated amino acids: Influence on IRMPD spectra. *J. Phys. Chem. A* 117, 1181–1188. doi: 10.1021/jp305263b

Steill, J. D., Szczepanski, J., Oomens, J., Eyler, J. R., and Brajter-Toth, A. (2011). Structural characterization by infrared multiple photon dissociation spectroscopy of protonated gas-phase ions obtained by electrospray ionization of cysteine and dopamine. *Anal. Bioanal. Chem.* 399, 2463–2473. doi: 10.1007/s00216-010-4582-y

Storn, R., and Price, K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optimiz.* 11, 341–359. doi: 10.1023/A:1008202821328

Sunahori, F. X., Yang, G., Kitova, E. N., Klassen, J. S., and Xu, Y. (2013). Chirality recognition of the protonated serine dimer and octamer by infrared multiphoton dissociation spectroscopy. *Phys. Chem. Chem. Phys.* 15, 1873–1886. doi: 10.1039/C2CP43296J

Tian, Z., and Kass, S. R. (2009). Gas-Phase versus Liquid-Phase Structures by Electrospray Ionization Mass Spectrometry. *Angew. Chem. Int. Ed.* 48, 1321–1323. doi: 10.1002/anie.200805392

Tian, Z. X., and Kass, S. R. (2008). Does Electrospray ionization produce gas-phase or liquid-phase structures? *J. Am. Chem. Soc.* 130:10842. doi: 10.1021/ja802088u

Vehkamäki, H. (2006). *Classical Nucleation Theory in Multicomponent Systems.* Berlin; Heidelberg; New York, NY: Springer-Verlag.

Verkouteren, J. R., and Staymates, J. L. (2011). Reliability of ion mobility spectrometry for qualitative analysis of complex, multicomponent illicit drug samples. *Forensic Sci. Int.* 206, 190–196. doi: 10.1016/j.forsciint.2010.08.005

Viehland, L. A., and Mason, E. A. (1995). Transport properties of gaseous-ions over a wide energy-range.4 *Atomic Data Nuclear Data Tables* 60, 37–95. doi: 10.1006/adnd.1995.1004

Wales, D. J., and Doye, J. P. K. (1997). Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A* 101, 5111–5116. doi: 10.1021/jp970984n

Wales, D. J., Miller, M. A., and Walsh, T. R. (1998). Archetypal energy landscapes. *Nature* 394, 758–760. doi: 10.1038/29487

Wales, D. J., and Scheraga, H. A. (1999). Review: chemistry - global optimization of clusters, crystals, and biomolecules. *Science* 285, 1368–1372. doi: 10.1126/science.285.5432.1368

Wang, J., Wang, W., Kollman, P. A., and Case, D. A. (2006). Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* 25, 247–260. doi: 10.1016/j.jmgm.2005.12.005

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244. doi: 10.1080/01621459.1963.10500845

Weininger, D. (1988). SMILES, a chemical language and information system: 1: introduction to methodology and encoding rules. *J. Chem. Inform. Comp. Sci.* 28, 31–36. doi: 10.1021/ci00057a005

Wickelmaier, F. (2003). *An Introduction to MDS*. Aalborg: Aalborg Universitetsforlag.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.