CHEMICAL REVIEWS

pubs.acs.org/CR

Review

Quantum Chemistry Calculations for Metabolomics

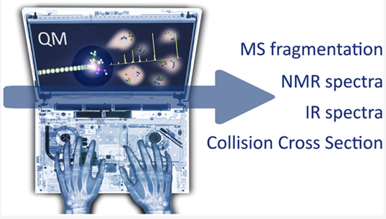
Focus Review

Ricardo M. Borges, Sean M. Colby, Susanta Das, Arthur S. Edison, Oliver Fiehn, Tobias Kind, Jesi Lee, Amy T. Merrill, Kenneth M. Merz, Jr., Thomas O. Metz, Jamie R. Nunez, Dean J. Tantillo, Lee-Ping Wang, Shunyang Wang, and Ryan S. Renslow*

Cite This: *Chem. Rev.* 2021, 121, 5633−5670

Read Online

ACCESS | Metrics & More | Article Recommendations



ABSTRACT: A primary goal of metabolomics studies is to fully characterize the small-molecule composition of complex biological and environmental samples. However, despite advances in analytical technologies over the past two decades, the majority of small molecules in complex samples are not readily identifiable due to the immense structural and chemical diversity present within the metabolome. Current gold-standard identification methods rely on reference libraries built using authentic chemical materials ("standards"), which are not available for most molecules. Computational quantum chemistry methods, which can be used to calculate chemical properties that are then measured by analytical platforms, offer an alternative route for building reference libraries, *i.e.*, *in silico* libraries for "standards-free" identification. In this review, we cover the major roadblocks currently facing metabolomics and discuss applications where quantum chemistry calculations offer a solution. Several successful examples for nuclear magnetic resonance spectroscopy, ion mobility spectrometry, infrared spectroscopy, and mass spectrometry methods are reviewed. Finally, we consider current best practices, sources of error, and provide an outlook for quantum chemistry calculations in metabolomics studies. We expect this review will inspire researchers in the field of small-molecule identification to accelerate adoption of *in silico* methods for generation of reference libraries and to add quantum chemistry calculations as another tool at their disposal to characterize complex samples.

CONTENTS

Received: August 24, 2020
Published: May 12, 2021

https://doi.org/10.1021/acs.chemrev.0c00901

1. INTRODUCTION

1.1. Growth and Impact of the Omics

Current basic and applied research of living systems occurs amid several rapidly evolving scientific paradigms, omics (*e.g.*, genomics, transcriptomics, proteomics, and metabolomics),[1−7] systems biology,[8−10] and synthetic biology,[11−13] that influence the researcher to look broadly at the holistic system or organism under study. Propelled by key developments of the Information Age, these scientific paradigms encourage scientists to aim for the comprehensive characterization and quantification of the relevant functional units of a cell, organ, organism, or entire system (*e.g.*, soil) and to develop computational models that capture and explain the interactions between and among these units that influence the overall system (Figure 1). At the lowest level, the units that comprise those systems are genes, transcripts, proteins, and metabolites; these units are responsible for the mechanisms by which interactions occur and lead to higher-level system functions and properties. Here, we use "metabolites" to refer not only to small molecules involved in primary metabolism but also to secondary (or "specialized") metabolites. Secondary metabolites are typically defined as molecules that are not directly involved in organism growth, development, or reproduction[14−16] but instead are produced as a consequence of interactions with other organisms and the environment (*e.g.*, signaling, defense/deterrence, and larger biomolecule degradation).[17−20] Related small molecules that are equally important include polar and nonpolar lipids and anthropogenic molecules, such as pesticides, fertilizers, and pharmaceutical products. Similarly, glycans are polysaccharide moieties often bound to proteins on cell surfaces important for cell recognition but also may be bound to lipids or occur freely after enzymatic release.

The measurement of each of the classes of biomolecules that comprise low-level functional units has been enabled by their respective omics paradigm, genomics, transcriptomics, proteomics, metabolomics, lipidomics, and glycomics (Figure 2), and the numbers of publications including data from such studies has steadily increased over the last two decades (Figure 3). Major funding agencies have increasingly recognized the high data yield of omics approaches and their potential to generate new biological and biomedical hypotheses. Many research studies today include one or more types of omics, and many research consortia, centers, and cores focus on multiomics approaches to studying health and disease or provide omics measurement services to clients. Indeed, the U.S. National Institutes of Health (NIH) committed >$200 M in 2019−2020 to fund 266 grants or subgrants that include some aspect of omics data collection or analysis in their proposed research (NIH RePORTER search; keyword "omic" and limited to project abstracts).

Omics studies to date have yielded important discoveries of the roles of functional biomolecules or of the genes and pathways that encode or regulate them. For example, in the early 1990s, genomics-based research led to the discovery of two cancer-susceptibility genes, *BRCA1* and *BRCA2*,[21−25] that have revolutionized breast cancer screening. Women who inherit certain mutations in *BRCA1* or *BRCA2* have 72% and 69% risk, respectively, of developing breast cancer by age 80.[26] Mutations in these genes also increase the risk of ovarian cancer.[26]

In 2004, Zhang and colleagues used a proteomics approach to analyze sera from over 500 individuals with various ovarian cancers and benign pelvic masses in a five-laboratory, case-control study and using a robust study design.[27] A number of candidate protein biomarkers were identified, immunoassays were developed for subsequent validation in independent cohorts, and the results indicated that the marker panel could discriminate between benign and malignant ovarian tumors. The research group then worked collaboratively with the U.S. Food and Drug Administration (FDA) to develop an approved assay, and the FDA provided clearance in 2009. The final assay, OVA1, provides >90% sensitivity and 90% specificity for women with an ovarian tumor and for whom surgery is planned when combined with other data. The OVA1 assay is now commercially available from ASPiRA Laboratories to detect ovarian cancer risk in women with planned surgery for a pelvic mass.

Perhaps the most important impact of metabolomics to date is the contribution of early generation approaches to the identification of and monitoring for inborn errors of metabolism,[30,31] which are typically characterized by accumulation of high levels of key metabolites in blood and urine of those afflicted. The first disease identified through newborn screening was phenylketonuria (PKU), which is diagnosed in part based on high levels of circulating phenylalanine due to mutations in the gene encoding the hepatic enzyme phenylalanine hydroxylase.[32] If undetected or left untreated, PKU can lead to significant intellectual disability, among other ailments.[33] Today, every state in the U.S. supports screening programs for a wide range of inborn errors of metabolism; for example, Washington State currently tests for 36 disorders, and California state law now requires screening for 80 congenital and genetic disorders in all newborns.

1.2. High Throughput Omics Measurements

The analytical tools used for comprehensive omics measurements vary according to the biochemical nature of the molecules involved. The foundational knowledge of the chemical composition and molecular structure of DNA[34−36] and the molecular biology associated with the molecule[37−39] are the primary elements that have enabled present-day technology for rapid, comprehensive, and cost-effective determination of DNA composition and order ("sequencing"). DNA is a relatively simple biomolecule, consisting of two complementary, polymeric strands comprising repeating units of just four nucleotide bases. A variety of next-generation sequencing technologies are available today; however, at the

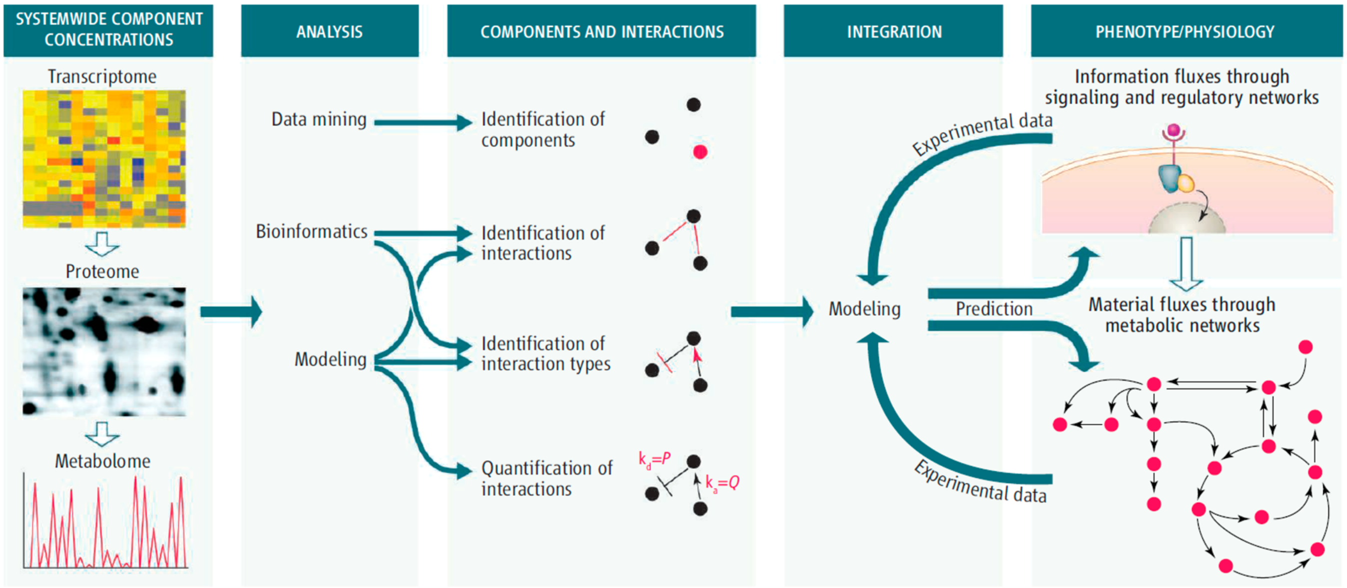

SYSTEMWIDE COMPONENT CONCENTRATIONS
ANALYSIS
COMPONENTS AND INTERACTIONS
INTEGRATION
PHENOTYPE/PHYSIOLOGY
Transcriptome
Proteome
Metabolome
Data mining
Bioinformatics
Modeling
Identification of components
Identification of interactions
Identification of interaction types
Quantification of interactions
$k_d=P$
$k_a=Q$
Experimental data
Modeling
Prediction
Experimental data
Information fluxes through signaling and regulatory networks
Material fluxes through metabolic networks

Figure 1. Systems biology paradigm. Systems biology studies employ omics approaches to comprehensively identify and quantify the functional units of the system under study. One or more omics approach is used to perform measurements of genes, transcripts, proteins, and metabolites, the data are analyzed and integrated, and computational models are used to interpret the results, often with the goal of obtaining a predictive understanding of the system to then manipulate it in a directed fashion. Reproduced with permission from ref 8. Copyright 2007 AAAS.

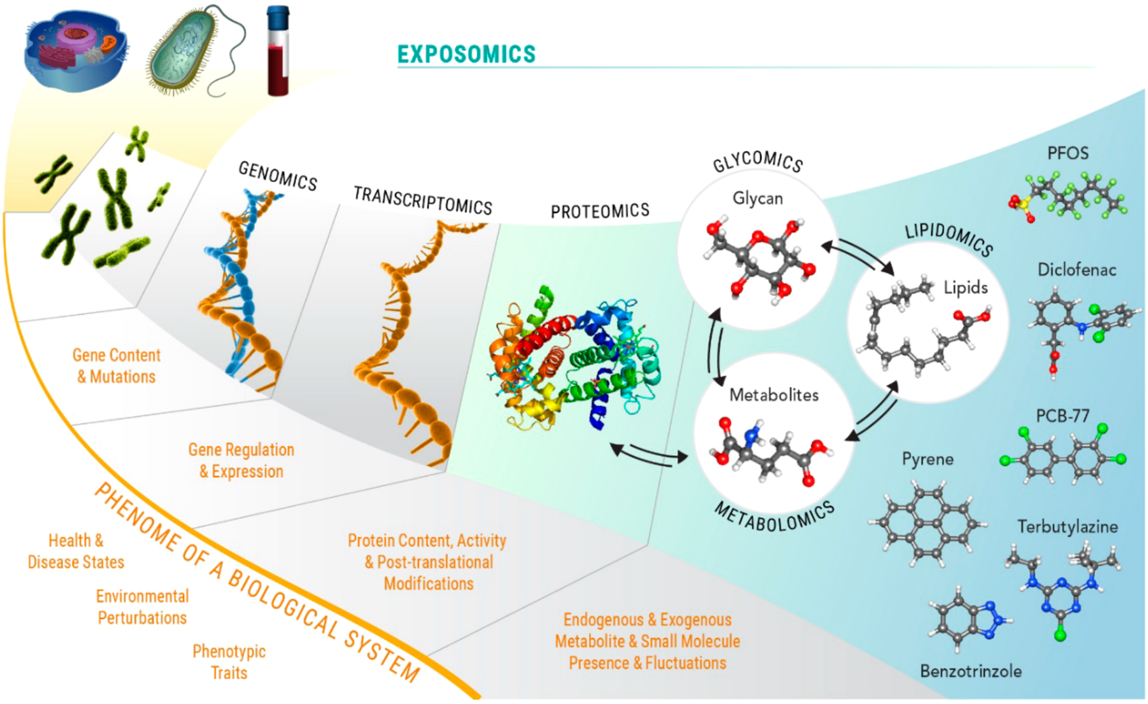

EXPOSOMICS
GENOMICS
TRANSCRIPTOMICS
PROTEOMICS
GLYCOMICS
Glycan
LIPIDOMICS
Lipids
Metabolites
METABOLOMICS
PFOS
Diclofenac
PCB-77
Pyrene
Terbutylazine
Benzotrinzole
Gene Content & Mutations
Gene Regulation & Expression
Protein Content, Activity & Post-translational Modifications
Endogenous & Exogenous Metabolite & Small Molecule Presence & Fluctuations
Health & Disease States
Environmental Perturbations
Phenotypic Traits
PHENOME OF A BIOLOGICAL SYSTEM

Figure 2. Omics. The approaches (and philosophies) for comprehensively identifying and quantifying genes, transcripts, proteins, and metabolites are termed genomics, transcriptomics, proteomics, and metabolomics, respectively. Lipidomics is the subdiscipline of metabolomics that addresses the measurement of polar and nonpolar lipids. Glycomics is the omics devoted to the comprehensive measurement of free and protein-bound glycans (as well as glycolipids, *i.e.*, lipid-bound glycans). The exposome includes all endogenous and exogenous exposures and unites transcriptomics, proteomics, metabolomics, lipidomics, and glycomics and includes measurement of anthropogenic molecules. Modified with permission from ref 28 under the Attribution-NonCommercial-No Derivatives 4.0 Unported License (http://creativecommons.org/licenses/by-nc-nd/4.0). Modified with permission from ref 29. Copyright 2016 Springer Nature.

highest level, all genomic sequencing begins with isolation of DNA from a sample, the shearing of the double-stranded molecule to a single strand, and the subsequent elongation of a short complementary primer sequence through sequential addition of free nucleotides by the action of the enzyme DNA polymerase.[3] The sequential incorporation of free nucleotides into the growing DNA chain is monitored by fluorescence detection of fluorophores bound to the nucleotides. The inherent specificities of DNA polymerase, hydrogen bonding between complementary pyrimidine and purine nucleotides,

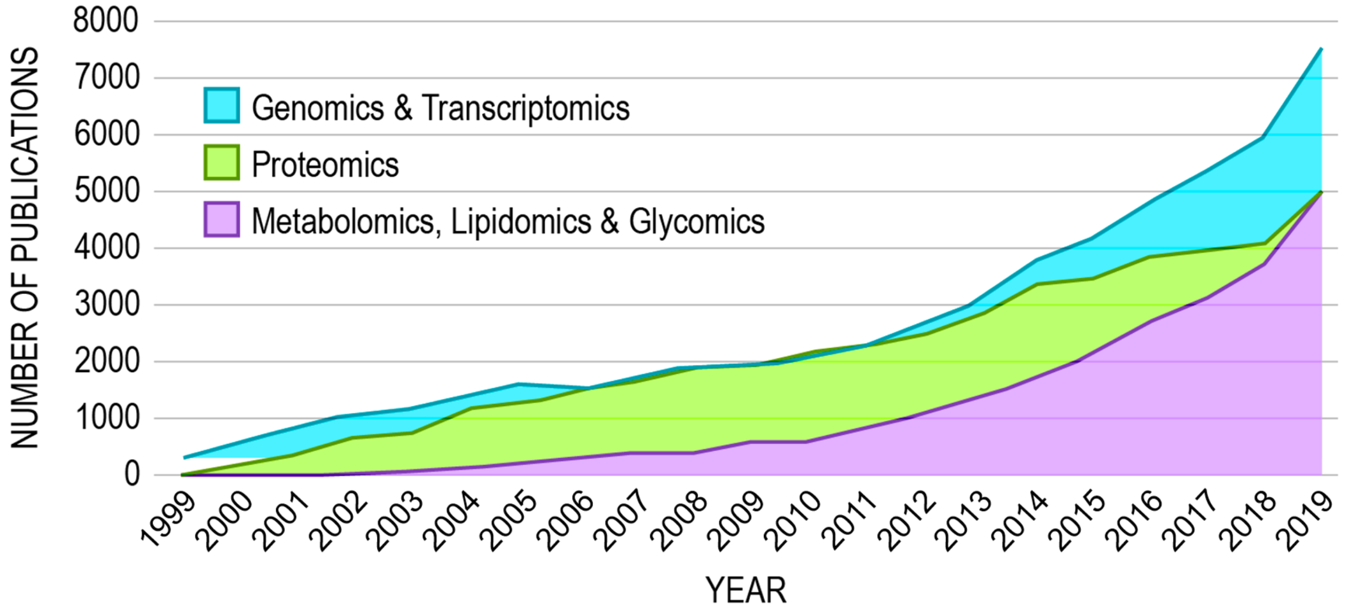


Figure 3. Omics publication trends 1999−2019. The numbers of publications including genomics, transcriptomics, proteomics, metabolomics, lipidomics, and glycomics approaches have steadily increased in the last two decades, linearly from 1999 to 2009 and exponentially thereafter. Results culminated from PubMed keyword searches of "genomics," "transcriptomics," "proteomics," "metabolomics" (and "metabonomics"), "lipidomics," and "glycomics" and limited to appearance in publication title or abstract. Genomics and transcriptomics publications are combined because these approaches rely on sequencing technologies. Metabolomics, lipidomics, and glycomics publications are also combined because comprehensive analysis of these molecular types typically rely on mass spectrometry (MS) approaches but often involve other technology (such as nuclear magnetic resonance spectroscopy).

and nucleotide-bound fluorophores, combined with the accuracy of current sequencing data processing algorithms, all contribute to genomic sequencing results with very low error rates (typically less than 1%). The massive parallelization available in modern sequencing instruments allows for nearly complete coverage of a genome in a relatively short time and at low cost. Technologies for sequencing of RNA (*i.e.*, transcriptomics) are similar.

Proteins, like DNA and RNA, are polymers of repeating units of 20 amino acids. Unlike DNA and RNA, no molecular biology can be leveraged to determine their sequence in such a complete, accurate, and cost-effective manner. Instead, proteins are normally "sequenced" in proteomics analyses using tandem mass spectrometry (MS/MS). In the shotgun proteomics paradigm, proteins are digested into their constituent peptides using the enzyme trypsin, which cleaves on the carboxyl side of arginine and lysine residues. This process generates peptides of manageable size that are amenable to separation using liquid chromatography (LC), ionization using electrospray (ESI), and gas-phase fragmentation using, for example, collision-induced dissociation (CID).[1] During CID, peptides typically dissociate at the peptide bond, producing fragmentation spectra with constituent m/z corresponding to different sizes of the peptide in question, minus one or more constituent amino acids in sequence (*i.e.*, so-called "ion ladders"). Various software tools have been developed for *in silico* prediction of peptide fragmentation spectra.[40,41] These software tools essentially generate comprehensive reference libraries of predicted peptide spectra and for every protein suspected of being present in the sample. The predicted fragmentation spectra generated by these algorithms are based on peptides derived *in silico* from reference protein sequences, and the reference protein sequences are in turn generated from the genome of an organism of interest, thus showing the intimate relationship between genomics and proteomics. Because of comparatively higher errors in peptide identification using MS-based proteomics (compared to gene sequencing), approaches were developed to provide a measure of confidence in the results from proteomics data processing tools.[42] The most commonly implemented approach for estimating and controlling the error rate in proteomics data processing is the target−decoy database approach,[43] which allows researchers to control the degree of false identifications (*i.e.*, the false discovery rate) by setting minimum peptide identification score thresholds that both maximize the numbers of confidently identified peptides while minimizing the numbers of incorrect identifications.

1.3. Major Roadblocks in Metabolomics

Metabolomics is the least mature of the omics. Although the average molecular formula composition of a metabolite does not differ significantly from that of a peptide,[44] metabolite structures are not constrained to a template like DNA, RNA, and proteins. Their chemical diversity is governed only by what constitutes a thermodynamically stable structure (>10^{33}−10^{160} possible structures for molecules, depending on the number of atoms and elements considered[45,46]). Moreover, the concentrations of metabolites vary by over 10 orders of magnitude. Because of this immense chemical diversity and consequent broad range of physicochemical properties and abundances, multiple analytical technologies are employed in comprehensive metabolomics studies to achieve high coverage of the metabolome. The chemical diversity of the metabolome has also precluded the development of analytical paradigms providing high-throughput (*i.e.*, automated) and accurate identifications of metabolites with associated estimates of false discovery.

For novel molecules, using advanced ultrahigh resolution MS, the chemical formula can be readily determined with high confidence, but the organization of the constituent atoms into chemical structures cannot be unambiguously determined from many possible isomeric compounds with the same molecular formula.[47] Kind and Fiehn reported that for the molecular formula $C_{15}H_{12}O_7$, 181, 166, and 129 matches were identified in a search of the chemicals in the Chemical Abstracts, Beilstein, and Natural Products databases, respectively.[47] More broadly, within 540 000 molecules selected from the Human Metabolome Database (HMDB)[48] and the DSSTox database,[49] 20% of the molecular formulas match more than five compounds, and 474 000 molecules have a formula conflict with at least one other molecule. Thus, even with other properties such as isotopic signature, chemical structures cannot be unambiguously identified[28,47,50] for novel molecules from mass alone without use of orthogonal analysis (*e.g.*, chromatography, ion mobility, MS/MS) and comparison of experimental data to that from analyses of authentic reference chemicals. This problem has been especially prevalent in spatially resolved metabolomics (*i.e.*, imaging) applications, where, until recently, it has been challenging to add orthogonal dimensions of data for improved molecular identification.[51−54] Nuclear magnetic resonance (NMR) spectroscopy is an established tool for assignment of chemical structures to novel molecules but requires higher sample concentration and purity, limiting its utility for structural elucidation of novel molecules in a high throughput, comprehensive manner. However, small-volume NMR probes[55−57] at high field strengths are greatly improving the sensitivity limitations, and several approaches of mixture analysis are reducing or eliminating the need for purity.[58−60] Likewise, microcrystal electron diffraction (MicroED) has recently been demonstrated for direct and confident structure confirmation.[61]

For measuring known molecules, efficient analytical methodologies for confident identification of large numbers of metabolites in high throughput metabolomics studies are gas chromatography−MS (GC-MS), LC-MS, and NMR. Here, metabolite identification is achieved by comparison of experimental data to reference libraries containing data from analyses of authentic chemical standards. Such approaches satisfy the recommendations of the Metabolomics Standards Initiative (MSI) of the Metabolomics Society for confident metabolite identification.[62,63] However, the reliance of these approaches on reference data generated through analyses of authentic chemical standards is a significant limitation because the number of chemicals available for purchase is very limited relative to the number of molecules proposed to exist in the universe.[64] For example, HMDB represents <5% of the estimated total metabolite space across multiple organisms, and only ∼10% of HMDB molecules are represented by readily available authentic chemical standards[65] (verified through custom Python scripts to search known vendors).[7,66] Further, one of the largest repositories of authentic reference spectra, the Wiley Registry, contains data for nearly 300 000 molecules, or just <1% of known chemicals when considering the ChemSpider, PubChem, and American Chemical Society's CAS databases, which contain entries for tens of millions of chemicals.[67]

There is therefore a tremendous disparity between the numbers of metabolites, exposure molecules, and xenometabolites that can be confidently identified in metabolomics studies when adhering to current MSI guidelines *versus* the number of molecules postulated to fill "chemical space." A reasonable approach to increasing the amount of reference data for use in small-molecule identification is through *in silico* means. This review will discuss the potential for quantum chemistry approaches to contribute to the calculation of, for example, chemical properties and reference spectra for metabolites and other small chemicals, which can be used to aid molecular identification in complex samples, thereby overcoming a significant obstacle remaining in the field of metabolomics.

1.4. Quantum Chemical Applications

1.4.1. Historical Overview. The purpose of this section is to illustrate the promise of quantum chemistry for metabolomics, including prediction of quantities relevant to NMR, MS, and other methods. While we summarize some approaches here, numerous reviews and books are available that cover the applications of quantum chemistry in various subfields in greater detail.

Quantum chemistry is concerned with calculating the states and properties of the electrons in a molecular system using the laws of quantum mechanics (QM). According to the time-independent Schrödinger equation, the stationary quantum states of a system (*i.e.*, those with definite energy) are eigenfunctions of a Hamiltonian operator representing the electron kinetic energy and interactions among electrons and nuclei. This review will mostly limit itself to a conceptual framework bound by several assumptions such as the separation of nuclear kinetic and electronic energy scales (the Born−Oppenheimer approximation), nonrelativistic electronic energy scales, and the absence of fine structure-producing effects such as spin−orbit coupling. These assumptions are often made for computational efficiency without significantly sacrificing accuracy for many applications (including most presented here), although care must be taken when simulating systems where these effects are important, such as nonradiative relaxation around a conical intersection[68] or spin crossover induced by heavier elements.[69]

Schrödinger's equation does not have an analytic solution for the general many-electron problem. Exact numerical solutions are computationally inaccessible for all but the smallest systems. Thus, quantum chemistry involves finding approximate solutions that give the optimal compromise between accuracy and computational cost for problems of chemical interest. The Hartree−Fock theory[70,71] and density functional theory (DFT),[72−74] two of the very first quantum chemistry methods, were conceived in the 1920s while QM was still in its infancy. Significant advances were made in the 1960s and 1970s, partially motivated by the revolutions in technology and exponential increases in processing power that continued through the early 2000s. In 1998, John Pople and Walter Kohn were recognized with the Nobel Prize in Chemistry for their pioneering contributions in molecular quantum chemistry,[75,76] and one of its principal branches, Kohn−Sham DFT (KS-DFT).[77,78] The relationship of quantum chemistry to other molecular simulation approaches is illustrated in Figure 4.

Today, quantum chemistry is a flourishing field and continues to make significant advances both in terms of theoretical methods that afford increasingly accurate and efficient approximations and computational methods that take maximum advantage of available computer hardware and software libraries. It has become common practice for leading research groups in this field to release free or commercial software packages that implement quantum chemistry methods

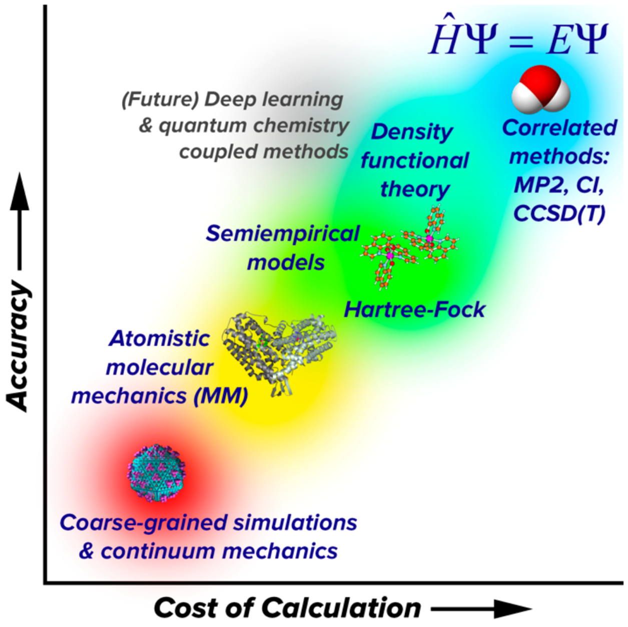

$\hat{H}\Psi = E\Psi$
(Future) Deep learning & quantum chemistry coupled methods
Density functional theory
Correlated methods: MP2, CI, CCSD(T)
Semiempirical models
Hartree-Fock
Atomistic molecular mechanics (MM)
Coarse-grained simulations & continuum mechanics
Accuracy
Cost of Calculation

Figure 4. Quantum chemistry methods (upper right) are considered highly accurate but also highly expensive compared to empirical potential-based simulation methods (lower left). Methods that combine physics-based principles with empirical knowledge, such as semiempirical models, density functional theory, and future deep-learning-based methods are promising for improving accuracy without increasing computational cost.

for the broader community to apply to chemical problems. Examples of these software packages include Gaussian,[79] Q-Chem,[80] GAMESS,[81] Psi4,[82] Molpro,[83] NWChem,[84] ORCA,[85] and TeraChem.[86,87]

Quantum chemistry has made significant impacts in the chemical sciences due to its ability to routinely compute many properties of experimental interest with "chemical accuracy", *i.e.*, accurate enough to make meaningful interpretations and predictions.[88] The usual working definition of chemical accuracy is that relative energies between two states should have an error of <1 kcal/mol compared to a gold standard, which can be experimental thermochemical data, such as atomization energies,[89,90] or another higher-level calculation. A commonly accepted gold standard calculation is coupled cluster with singles, doubles, and perturbative triples (CCSD(T)/CBS)[91,92] in the complete basis set limit.[93,94] In modern benchmarks, these values are computed indirectly using composite methods such as the Weizmann-*n* approaches that combine the results of several other calculations.[95] This approach is not a gold standard for all properties and systems because CCSD(T)/CBS is not a valid approximation for systems with significant multireference electronic character,[96−98] such as open-shell transition metal complexes or highly excited electronic states. Moreover, most experiments measure signals from a thermodynamic ensemble of molecules undergoing time evolution and sometimes in a condensed-phase environment, whereas an individual quantum chemistry calculation is carried out on a single molecular structure, which makes comparisons difficult. Simulating experimental observables using quantum chemistry often requires incorporating molecular dynamics (MD) on the quantum chemical potential energy surface (PES), configurational sampling, environmental effects, nuclear quantum effects, and possibly other effects if experimentally relevant. Thus, the definition of accuracy can vary broadly depending on the system and property being considered.

For many methods, it is possible to compute the analytic nuclear gradient of the energy at relatively low additional cost, which enables an efficient optimization of energy-minimized structures and other transition states on the molecular PES.[99,100] This affords a route to predicting the reaction energy and activation energy of hypothetical reaction mechanisms, providing important support for mechanistic understanding of chemical reactivity that is difficult to probe experimentally. A wide range of properties may be computed for a given molecular structure, including electronic and vibrational transition energies, electrostatic moments, and polarizability.[101] Calculations of nuclear shielding and internuclear couplings provide a route toward computing NMR observables (section 2.1). Still other properties may be computed from approximate free energy differences, such as redox potential[102] and pK_a.[103,104]

Many experimentally measured properties are derived not from a single structure but from a statistical ensemble at finite temperature, which could be simulated implicitly by making a rigid rotor/harmonic oscillator approximation[105,106] or sampled explicitly using MD[107,108] or Monte Carlo (MC)[109−111] methods. Effects of the solvent or other chemical environments are treatable implicitly by using polarizable continuum models[112] or explicitly by including molecules of the environment in the calculation, although the latter greatly increases the computational cost and dimensionality of configuration space. The quantum behavior of nuclei manifests as zero-point vibrational energy in the harmonic approximation or can be explicitly simulated using path-integral MD methods.[113−115] Hybrid models such as QM/MM are useful for explicitly modeling portions of the system, such as a solvent or protein environment, using inexpensive force field (FF) models.[116,117]

The applicability of quantum chemistry should be considered along several dimensions, including the size of the system, which affects the computational cost, the level of theory and amount of sampling needed to compute a result to desired accuracy and statistical precision, the availability of experimental data to inform the development of better methods and models, the importance and complexity of environmental effects, and last, the accessibility of the computational methods to nonexpert users.

1.4.2. Relative Energies and Equilibrium Structures. A basic building block of quantum chemistry applications is the calculation of relative energy between two chemical states, which may differ in the number of electrons, atoms, and three-dimensional structure. Several classes of relative energies include:

- Ionization energies and electron affinities[118,119]
- Proton affinities[120,121]
- Bond dissociation energies[122,123]
- Atomization energies, the total separation of molecules into constituent atoms
- Isomerization energies[124−126]
- Noncovalent interactions[127−129]
- Conformational relative energies[130−132]
- Reaction energies and activation energies (covered in detail in section 1.4.3)

In all of these cases, the calculation involves taking the difference between two calculated energies, often with the preliminary step of energy minimization. Numerous published benchmark studies describe the accuracy of various quantum chemistry methods *versus* the gold standard for these properties.[101,120,133−136]

Relative energies form the foundation of simulations that incorporate more aspects of the experimental system such as the thermodynamic ensemble, time evolution, environment, and nuclear quantum effects. It is important to note that to describe a chemical phenomenon, relative *free* energies are more meaningful and analogous to experimentally measurable quantities such as redox potential, pK_a, standard reaction free energies/free energies of activation, and equilibrium conformations.

In metabolomics and compound identification applications, the accurate prediction of relative peak heights is of principal importance.[137−139] Because the relative peak heights of adduct ions, *i.e.*, $[M + H]^+$, $[M + Na]^+$, *etc.*, vary depending on the analyte and solvent conditions, these data can be used in compound identification as discussed in greater detail below.[140] Recent developments in DFT functionals have seen nearly universal incorporation of empirical dispersion corrections[141−143] and consequently a dramatic improvement in accuracy for noncovalent interactions.[129,134,135] Therefore, one potential application of quantum chemistry is to predict the relative peak heights of adduct ions for an analyte by comparison of noncovalent interaction energies.[144,145]

1.4.3. Reaction Mechanism Analysis. Perhaps the most widespread application of modern quantum chemistry is the analysis of reaction mechanisms, which may be proposed from chemical intuition or by automated approaches. The basic computational results in these studies are reaction energies and activation barriers for elementary steps, which provide insight into the thermodynamic and kinetic feasibility of the proposed mechanism.[146−150] The reaction rate is proportional to $\exp[-E_a/RT]$, where E_a is the activation energy calculated from the potential difference between optimized transition state and reactant structures, R is the thermodynamic gas constant, and T is absolute temperature. The transition state is located at a saddle point on the PES, where the direction of downward curvature corresponds to a vibrational mode of imaginary frequency that leads toward the reactant and product structures on either side. At room temperature, an increase in ∼1.4 kcal/mol of E_a results in a roughly 1 order of magnitude decrease in reaction rate. Calculated E_a values are often used as a screen for room temperature mechanistic feasibility by comparison with a "rule of thumb" value, which typically ranges from 21 to 30 kcal/mol,[151−154] although values have been reported as high as 40 kcal/mol.[155]

These calculations neglect the contribution of system-dependent dynamical properties to the reaction rate but are still highly useful for mechanistic screening because computed values of E_a for different mechanistic hypotheses have a wide dynamic range of 0−100 kcal/mol or more. Multiple hypotheses may be compared under the assumption that differences in E_a play the dominant role in the relative rates. Care must be taken to select an appropriate method, as detailed benchmark studies in the literature have shown that many choices of quantum chemistry method and/or basis set could lead to errors of 10 kcal/mol or more.[129,135] Recently developed density functional approximations, such as range-separated hybrid functionals with dispersion correction (*e.g.*, ωB97X-V,[156] M08-HX-D3[135,157]) and double-hybrid functionals, are able to achieve accuracy to within 1−3 kcal/mol of the gold standard provided that rather large basis sets, such as def2-TZVPD, are used.[158] However, these methods may be cost prohibitive for larger systems, and all single-reference methods (including most DFT functionals) are suspect for systems that contain strong multireference electronic character. Free energy corrections are also important in reaction mechanism analysis; in particular, the inclusion of translational and rotational entropic contributions has a significant effect on association and dissociation reactions.[85]

A long-term goal of reaction mechanism analysis is to leverage mechanistic insights toward reaction and catalyst design. For example, computations may be used to choose the best candidate from a series of catalysts that follow the same mechanism but differ in their activation energies. Although quantum chemistry is starting to play a more active role alongside experiment in this arena, mechanistic investigations still require significant human and computational effort, and further developments are needed to make truly novel predictions and designs. Automated tools for mechanism generation are a recent development that show promise for making progress in this field.[159−166] Reaction mechanism analysis is important for metabolomics in terms of predicting fragmentation patterns in MS experiments, described later in this article.

1.4.4. Thermodynamic Properties. Quantum chemistry can provide insight into equilibrium properties of a system by accounting for the distribution of states in a thermodynamic ensemble. This is commonly done using simple models; one common approach is to model the molecular partition function using a product of ideal gas, rigid rotor, and harmonic oscillator terms, although several more advanced methods that go beyond this approximation are available.[167] The molecular vibrational frequencies are obtained from the Hessian (second nuclear derivatives) of the electronic energy. Alternatively, the thermodynamic ensemble can be sampled explicitly using MD or MC methods.

Some of the most important experimentally measurable quantities are functions of the free energy difference between two states. These quantities include the standard reduction potential (redox potential) and pK_a, which measure the tendency for a species to gain electrons and protons in solution, respectively. The accuracy of DFT in calculating redox potentials is well established, and it has been hailed as a tool for rational electrocatalyst design.[102,168] Errors in computed values *vs* experiment can be as small as 0.1−0.2 V, which is small compared to experimental variations across a typical series of redox-active compounds.

The predictions of redox potentials and pK_a values have an important dependence on the choice of solvent model. Widely used implicit solvent models, which are based on a polarizable continuum,[169−171] tend to have lowered accuracy in systems where solute−solvent hydrogen bonding effects are important. Perhaps owing to this difficulty, the sizes of errors in redox potential calculations tend to be more similar within a group of chemically similar compounds but vary more broadly across disparate compounds.[172] Errors incurred by implicit solvent are especially onerous for pK_a calculations, where the addition or removal of a proton often leads to differences in hydrogen bonding interactions. An error in the calculated ΔG of just 1.4 kcal/mol corresponds to a deviation of 1 pK_a unit from experiment.[173,174] The use of explicit solvent models has been shown to yield improvements in the accuracy of pK_a and redox potential calculations,[160,175] but they are much more computationally expensive, requiring significant amounts of sampling of solute and solvent degrees of freedom.

MD and MC are simulation methods that can sample from a thermodynamic ensemble of states. MD propagates the atoms in the system using classical equations of motion, and thermodynamic sampling is achieved by means of a thermostat that perturbs the molecular velocities in order to sample from the constant temperature ensemble. A barostat can similarly be used to perturb simulation cell volumes to sample from the constant pressure ensemble. In principle, MD can be used with any underlying potential function, and it is termed *ab initio* molecular dynamics (AIMD) when the potential is calculated using quantum chemistry methods. MC, on the other hand, performs sampling by making randomized proposed moves through configuration space with acceptance probabilities calculated from ratios of Boltzmann factors. Both methods have been applied to compute equilibrium properties of molecular and solid substances.

For ordinary substances such as liquid water, good agreement with experimental data can be reached for properties such as the bulk density and radial distribution function.[176,177] However, nuclear quantum effects need to be explicitly accounted for using path integral MD[113−115] or otherwise roughly approximated by increasing the simulation temperature. Because *ab initio* methods are not parametrized to experimental data, they have also been applied to predict new phases of matter at extreme conditions, including superionic phases of ice[178] and metallic phases of hydrogen,[109] which are hypothesized to exist in planetary cores but have yet to be confirmed experimentally.

In metabolomics applications, compounds are often separated based on differences in the strength of intermolecular interactions with immiscible solvent phases.[179−181] The key physical quantities that predict retention time are the octanol−water partition coefficient (log P) or distribution coefficient (log D), which is proportional to the difference in solvation free energies in the two phases.[182] Although empirical models such as neural networks have long been the dominant method for predicting partition coefficients,[183−185] we expect that quantum chemical methods will play an increasingly important role in predicting this important property in the future as sampling methods and computational efficiency continue to improve.

1.4.5. Spectroscopic and Excited State Properties.

A variety of quantum chemistry methods are available to model how molecules respond to electromagnetic radiation in a variety of energy regimes. These methods are broadly useful for interpreting and assigning spectra, including in compound identification applications. At the low end of the energy spectrum, NMR shielding tensors and internuclear couplings may be computed using DFT (section 2.1).

Infrared (IR) spectra, which probe molecular vibrations, can be calculated from quantum chemistry in several ways. One type of approach is based on expanding the PES around a minimum energy structure, starting with normal modes obtained from diagonalizing the Hessian matrix. Because this approach ignores higher-order anharmonic effects, method-dependent empirical scaling factors are applied to calculated harmonic vibrational frequencies in order to obtain improved agreement with experiment.[186] Going beyond the harmonic approximation, a fourth-order expansion called a quartic FF[187] can be used in conjunction with vibrational perturbation theory or vibrational configuration interaction to obtain rovibrational spectra including anharmonic effects.[188,189] In the other kind of approach, MD simulations can also be used to simulate IR spectra by taking the Fourier transform of the dipole autocorrelation function.[190,191] This method is useful for obtaining spectra in bulk liquids and macromolecules, where soft degrees of freedom and multiple minima are more prevalent. The application of IR to small molecule identification is discussed in section 2.3.

Visible and ultraviolet (UV) wavelengths induce electronic excitations, which require quantum chemical descriptions of the excited state. Building off of the foundations of Hartree−Fock and DFT, the configuration interaction singles (CIS)[192] and time-dependent DFT (TD-DFT)[193] methods build a Hamiltonian matrix in the space of single excitations from the reference ground-state wave function, followed by diagonalization to obtain excited state energies. TD-DFT tends to yield root mean-squared errors of ~0.3 eV in vertical excitation energies for many functionals,[133] but the topography of the excited state potential energy surface is incorrect, leading to incorrect descriptions of conical intersections where the excited-state and ground-state energies become degenerate. Multireference wave function methods[194,195] can yield improved accuracy for excited-state potential energy surfaces, which may be applied to light-activated molecular switches and other electronically excited molecules.

The accuracy of the equilibrium structure is usually considered to be secondary to the reproduction of relative energies in most applications. One important exception is in microwave (rotational) spectroscopy experiments that can provide very accurate values of the moments of inertia of gas-phase molecules; these are sensitive to changes in bond lengths to within 10^{-13} m and bond angles to within 0.01°.[122] The most accurate measurements of molecular geometry are derived from these experiments, and highly accurate optimized structures can in turn be used to assign microwave spectra of unknown compounds.[123]

2. CURRENT APPLICATIONS AND REAL EXAMPLES

2.1. Nuclear Magnetic Resonance

The main alternative to the MS-based approach to detect metabolites is provided by analytical NMR spectroscopy.[65,196−199] High-resolution NMR analysis is capable of providing accurate structures of a range of molecules including metabolites.[200] This is of import in metabolomics, where ultrasensitive mass spectrometry instruments can detect differential mass signals but do not provide enough structural information to structurally characterize a given metabolite.

To bridge the gap between spectroscopic observations and structure the metabolomics field has been turning to the tools of computational chemistry.[201−208] Over the past decade, *in silico* calculations of NMR chemical shifts have significantly improved in accuracy, affordability, and reliability. The overall improvements in NMR computation come from methodological advancements, increased computational power, as well as complete end-to-end automation of these calculations.[209,210] Therefore, reliable chemical shift calculations are now highly accessible to chemists.

Extensive literature exists in the field of NMR chemical shift computation. For example, Kaupp, Buhl, and Malkin edited a book on the calculation of NMR and electron paramagnetic resonance (EPR) parameters.[211] In 2008, Casabianca and de Dios published an extensive review on *ab initio* NMR chemical shift calculation.[212] Bryce and Wasylishen wrote an extensive review, including an overview of the calculation of NMR parameters using *ab initio* methods.[213] Additionally, a review of chemical shielding calculations on proteins, peptides, and amino acids has been provided by Oldfield.[214] Several applications of chemical shift calculations are described by Facelli's review.[215] Gauss and Stanton added electron correlation to the computation of chemical shielding.[216] Tossell used a cluster model to compute shielding values for crystals,[217] and Sebastiani reviewed chemical shift calculations in condensed phases.[218] Benzi *et al.*[219] and Bagno *et al.*[220] included solvent effects in chemical shift calculations, and Oldfield included electrostatic effects on chemical shifts and applied this improved method to protein structure determination.[221] On the other hand, Hunter *et al.* discussed the use of semiempirical methods to evaluate chemical shifts[222] as did Merz and co-workers.[223] For rapid NMR shift computation for larger systems, a QM/MM approach has been described.[224,225] Finally, Bryce and Sward summarized the state-of-the-art for the study of quadrupolar halogens using solid-state NMR.[226]

There is an extensive literature of the use of quantum mechanical methods to explore NMR related problems, including the determination of relative stereochemistry[227−229] and NMR assignment of regioisomers.[230] Tantillo, Siebert, and Lodewyk review the use of *ab initio* and DFT calculations for NMR chemical shift prediction.[231,232] Benchmarking studies for the prediction of chemical shifts with different methods and levels of theory have been reported.[233−235] Further summaries on advancements in the evaluation of chemical shielding values can be found in Jameson and de Dios.[236,237]

To the best of our knowledge, the studies mentioned above only considered single conformer data sets for computation of NMR chemical shifts and accuracy assessment. In the case of flexible molecules with multiple conformers, the field is considerably less mature. For flexible molecules (*e.g.*, metabolites or drug-like molecules) with a large chemical space, considering a single conformer for these calculations most likely yields erroneous results.[238] Therefore, to get meaningful NMR data for flexible compounds, a large chemical space (*i.e.*, robust conformational search) needs to be considered. There have been three successful attempts to design an automatic protocol to predict NMR chemical shift of flexible molecules. Willoughby *et al.*, Yesiltepe *et al.*, and Grimme *et al.* have developed automated approaches for ^{1}H and ^{13}C chemical shift prediction for the nonexpert.[204,209,210] The computational protocol of Willoughby *et al.* is shown in Figure 5. Another protocol was developed by Xin *et al.* to compute ^{13}C NMR chemical shifts (Figure 6), employing the standard DFT method with an optimized basis set (cc-pVDZ) and a DFT functional (B3LYP) for organic molecules.[239] All of these protocols employ high-level QM methods, yielding accurate and reliable NMR results, but due to the expense of the QM methods used, they are too computationally expensive to handle a large number of samples with unknown structures. Moreover, more complex molecules like metabolites, with many rotatable bonds, offer challenges associated with sampling complex conformational spaces and the extant protocols have yet to be fully validated.

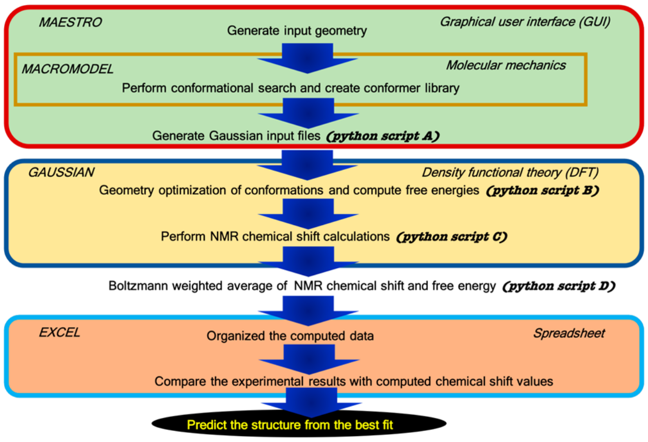

MAESTRO
Graphical user interface (GUI)
Generate input geometry
MACROMODEL
Molecular mechanics
Perform conformational search and create conformer library
Generate Gaussian input files (python script A)
GAUSSIAN
Density functional theory (DFT)
Geometry optimization of conformations and compute free energies (python script B)
Perform NMR chemical shift calculations (python script C)
Boltzmann weighted average of NMR chemical shift and free energy (python script D)
EXCEL
Organized the computed data
Spreadsheet
Compare the experimental results with computed chemical shift values
Predict the structure from the best fit

Figure 5. NMR chemical shift calculation protocol developed by Willoughby *et al.*[204]

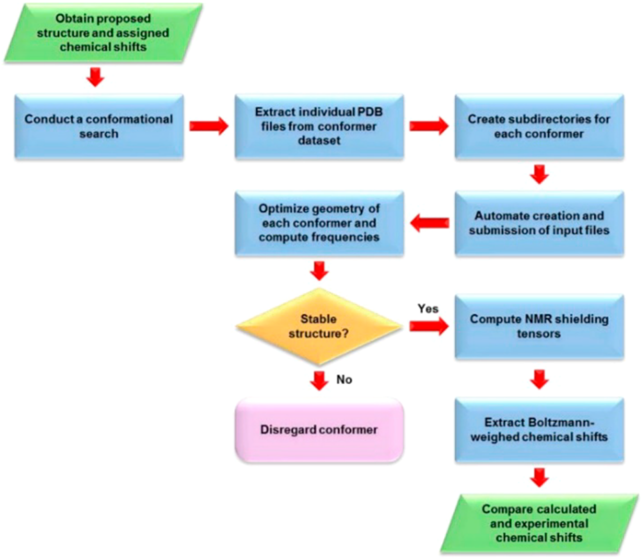

Obtain proposed structure and assigned chemical shifts
Conduct a conformational search
Extract individual PDB files from conformer dataset
Create subdirectories for each conformer
Automate creation and submission of input files
Optimize geometry of each conformer and compute frequencies
Stable structure?
Yes
No
Compute NMR shielding tensors
Disregard conformer
Extract Boltzmann-weighed chemical shifts
Compare calculated and experimental chemical shifts

Figure 6. NMR chemical shift calculation protocol developed by Xin *et al.*[239] Reproduced with permission from ref 239. Copyright 2017 American Chemical Society.

An efficient computational workflow must use the most computational efficient method at each step to accomplish the desired goal. Accurate QM methods are very computationally expensive and should only be used when no other options are available.[240,241] Where possible less expensive FF methods can be used, but attention should be paid to their overall accuracy. Machine learning (ML) methods can also be substituted where appropriate.[242,243] In the approach of Das *et al.*, their workflow takes advantage FF, ML, and QM-based methods to generate structural predictions for medium-size organic molecules including metabolites.[244] The pipeline encompasses the following steps (Figure 7): (i) conformation generation using an FF-based method, (ii) filtering the FF generated conformations using the Atomic Simulation Environment-Accurate Neural Network Engine for Molecular Energies (ASE-ANI) model, (iii) clustering of the optimized conformations based on structural similarity to identify chemically unique conformations, (iv) DFT structural optimization of the unique conformations, and (v) DFT NMR chemical shift calculation.[244] This protocol can calculate the NMR chemical shifts of a set of molecules using any available combination of DFT theory, solvent model, and NMR-active nuclei, using both user-selected reference compounds and/or linear regression methods. The protocol reduces the overall computational time by 2 orders of magnitude over methods that optimize the conformations using fully *ab initio* methods

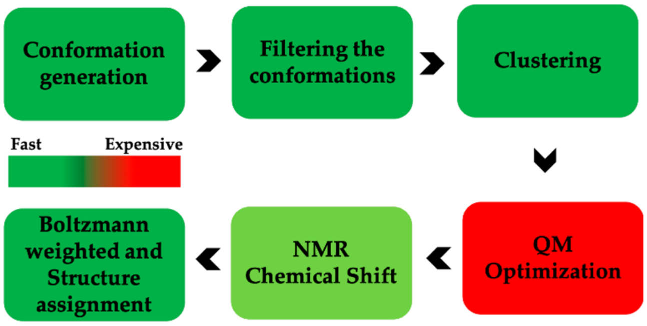
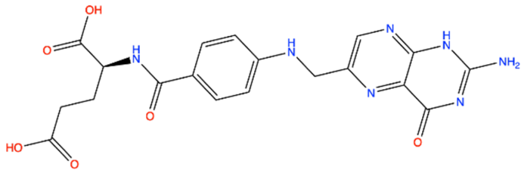


Figure 7. Das and Merz protocol to calculate NMR chemical shift. This protocol includes FF, ML-QM, and standard QM methods to improve efficiency of NMR computation technique with low computational cost.

yet matches experimental structural observations. The complete protocol provides an efficient way to obtain chemical shifts for conformationally flexible metabolites.

To illustrate the NMR chemical shift (^{1}H and ^{13}C) protocol of Das and co-workers, the case of folate ($C_{19}H_{19}N_7O_6$: Figure 8) is presented herein. Folate was run through the workflow illustrated in Figure 7. The detailed results are given in Figure 9 and in Tables 1−Table 3. Overall, the mean absolute error (MAE) values of ^{1}H and ^{13}C are 0.26 and 1.63, respectively, confirming a good agreement of computational data with the experimental chemical shift. MAE values are calculated using eq 1.

$$\text{MAE} = |\Delta\delta_{\text{av}}| = \frac{1}{N}\sum_{i=1}^{N} |\delta_i^{\text{comp}} - \delta_i^{\text{exp}}| \tag{1}$$



Metabolite	BMRB ID	No. of atoms	No. of rotatable bonds	Conf. Nos. (FF)	ANI Optimized Conf. Nos.	No. of clusters
Folate	bmse000299	51	10	561	561	26

Figure 8. BMRB ID, no. of atoms, no. of rotatable bonds, FF generated conformation, ANI optimized conformations, and no. of clusters are reported for folate molecule.

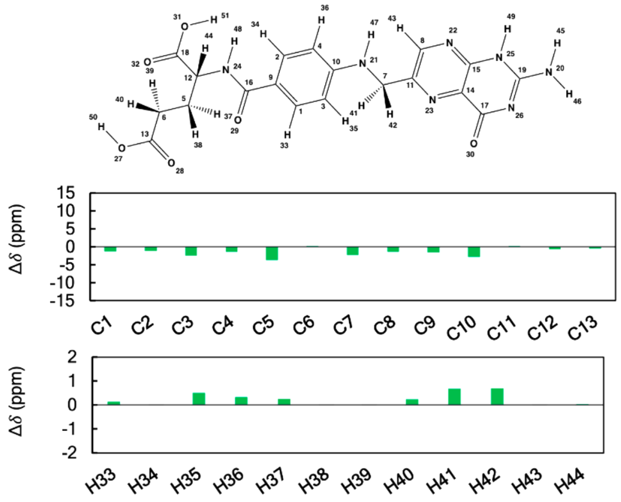


Figure 9. Plots of the differences between the calculated and experimental ^{1}H and ^{13}C NMR chemical shifts of folate. Shielding constants were computed at the B3LYP/6311G+(2d, p) level of theory and converted to linear scaled reference chemical shifts. Values of chemical shift differences are given in ppm.

Table 1. Relative Energies, Boltzmann Factor, and Equilibrium Mole Fraction of All Structurally Distinct Conformations of Folate

conformation no.	relative energy (kcal)	Boltzmann factor	equilibrium mole fraction
1	0.36	0.55	0.17
2	2.94	0.01	0.00
3	2.34	0.02	0.01
4	2.93	0.01	0.00
5	3.42	0.00	0.00
6	1.65	0.06	0.02
7	2.52	0.01	0.00
8	2.71	0.01	0.00
9	3.09	0.01	0.00
10	6.83	0.00	0.00
11	0.39	0.51	0.16
12	2.91	0.01	0.00
13	1.20	0.13	0.04
14	4.17	0.00	0.00
15	0.46	0.46	0.15
16	2.19	0.02	0.01
17	7.47	0.00	0.00
18	3.82	0.00	0.00
19	1.20	0.13	0.04
20	1.77	0.05	0.02
21	4.44	0.00	0.00
22	6.18	0.00	0.00
23	2.68	0.01	0.00
24	0.00	1.00	0.32
25	3.58	0.00	0.00
26	1.18	0.14	0.04

Table 2. Computed and Available Experimental ^{1}H NMR Shifts for Folate

atom no.	computed chemical shift (ppm)	experimental chemical shift (ppm)
H33	7.52	7.37
H34	7.39	7.37
H35	6.69	6.17
H36	6.51	6.17
H37	2.43	2.15
H38	2.07	2.05
H39	2.35	2.32
H40	2.57	2.32
H41	4.67	3.94
H42	4.66	3.94
H43	8.39	8.37
H44	4.34	4.27
H45	4.93	
H46	4.92	
H47	5.33	
H48	6.29	
H49	7.69	
H50	8.14	
H51	9.09	

Table 3. Computed and Available Experimental ^{13}C NMR Chemical Shifts for Folate

atom no.	computed chemical shift (ppm)	experimental chemical shift (ppm)
C1	129.92	131.32
C2	129.99	131.32
C3	111.41	113.93
C4	112.38	113.93
C5	27.20	31.09
C6	37.42	37.09
C7	45.05	47.45
C8	149.48	150.97
C9	150.86	152.47
C10	120.09	123.10
C11	150.75	150.41
C12	57.95	58.65
C13	184.59	185.20
C14	125.97	
C15	143.16	
C16	166.65	
C17	166.79	
C18	172.26	
C19	150.36	

where δ_i^{exp} is the experimental NMR chemical shift value of the ith nucleus of a molecule and δ_i^{comp} is the computed NMR chemical shift of same nucleus. This high-throughput workflow[244] can be deployed to obtain the chemical shifts for large collections of candidate metabolite structures to facilitate their characterization.

2.2. Ion Mobility Spectrometry

Collision cross section (CCS, units: Å^2) is a gas-phase property of a molecule that is obtained using ion mobility spectrometry (IMS). In IMS, ionized molecules enter a drift tube or other ion conduit that contains an electric field to accelerate sample ions and a buffer gas to produce a countering drag force, resulting in separation based on the ion's size, shape, mass, charge, conformational population, and interaction with the buffer gas.[245] IMS is the gas-phase separation method analogous to liquid phase separation by electrophoresis, which also relies on molecular interactions with a buffer and electrical mobility to provide a force for acceleration toward a detector. Depending on the instrument, CCS can be calculated from the time taken to drift (as in drift time IMS) before detection (typically at a mass spectrometer) or via calibration against analytes with known CCS (as in traveling wave IMS). While IMS-MS techniques have been explored for nearly six decades, the use of CCS as a complementary property to mass and chromatography-derived retention time for identification of small molecules has only become popular, by way of instrument commercialization,[246] in the past decade.[245] IMS does not rely on condensed phase interactions that can be subject to degradation, contamination (*e.g.*, heavy carryover), high variability, and difficult manufacturing requirements. Thus, measured CCS values can be very consistent over time and between laboratories, reaching relative standard deviations lower than 0.3%.[247] Furthermore, because CCS is obtained after ionization, each analyte in the sample of interest can be represented by multiple adduct ions, a function of the analyte—matrix interaction, each with its own CCS and mass. As a result, observing multiple adduct ions lends additional evidence to the presence of a particular molecule.[248]

To better understand molecular ion behavior in ion mobility instruments, the prediction of CCS from molecular ion structures bourgeoned in the 1990s.[249,250] It was not until the early 2000s that researchers began to routinely use DFT methods to generate gas-phase molecular ion structures as a foundation for CCS calculations.[251−253] Currently, DFT-based calculations are considered the gold standard method in determining the adduct conformations that give rise to measured CCS; however, predicting CCS for use in small-molecule identification studies is quickly shifting to machine and deep learning methods (discussed in section 4.2) due to their speed and accuracy.[254−258] Typical procedures using DFT for CCS calculations include the prediction of the bonding structure of the ionized adduct (*e.g.*, specific atom site of ionization), generation of a population of conformers of the adduct, optimization of conformer geometries, relative energy calculations for the conformers, CCS calculation for each conformer, and finally a method of either combining CCS values of conformers into a single combined CCS value for that conformational population (*e.g.*, averaged or Boltzmann-weighted) or selecting a conformer that is most likely to represent an experimentally observed CCS peak. DFT is typically used for the geometry optimization and energy calculation steps, but it can also be used for ionization site predictions and the generation of conformer populations.

Some of the earliest examples of DFT use in CCS calculations for ion mobility experiments came out of Prof. Michael Bowers' (University of California at Santa Barbara) and Prof. David Clemmer's (Indiana University) groups. In a study by Wyttenbach, Witt, and Bowers, DFT calculations were used for a set of conformers (generated using the Assisted Model Building with Energy Refinement, or AMBER, molecular dynamics suite) of glycine, alanine, and their methyl-substituted derivatives.[259] The study showed agreement between the predicted and observed CCS values, demonstrated the ability to determine likely conformations of the adducts and found differences in the adduct ion attachment (charge solvation, where the charge is stabilized by an electron dense region of the neutral molecule, and salt bridge, where the charge interacts with the zwitterionic form of the molecule).[259] In a study by Leavell *et al.*, diethylenetriamine—hexose complexes were investigated with CCS calculated using DFT-derived geometries.[252] Geometry optimization was performed using a low level of DFT theory along with a semiempirical method (AM1), with subsequent energy calculations performed at a higher level of theory. A small set of low energy candidate structures were found that match well to experimental CCS values, with evidence pointing to chair or twist-boat conformations for the hexose portion of the complexes.

While most early studies using DFT in CCS calculations focused on using these approaches to determine likely conformer structures of known molecules and complexes, it must be noted that even for low energy conformations, wildly different conformational structures can give rise to CCS values that are close enough to be experimentally indistinguishable with current instruments. Furthermore, CCS peaks of individual adducts can be much broader than the difference in peak max (center) CCS values between different molecules, reflecting the diverse conformational populations of ion packets. This issue becomes more pronounced with increased mass and decreased molecular rigidity (*i.e.*, as degrees of freedom increase). Because most metabolomics studies using

IMS operate at room temperature, adducts can rapidly interchange between conformations, as revealed by observed CCS peak distributions. Thus, we find the most value in DFT calculations for use in predicting CCS distributions to build identification libraries (*i.e.*, to build evidence for identification of the underlying adduct) as opposed to determining the underlying conformer of an observed IMS peak. For determining specific conformations of gas-phase ions (especially when the molecule of interest is already known and purposefully being studied), we recommend cryogenic IR ion spectrometry,[260] which is extremely sensitive to small changes in molecular conformation and provides tight constraints for calculations of molecular structures[261] (further details in section 2.3).

Through the pioneering work of Iain Campuzano and colleagues, the use of quantum chemical-based theoretical calculations of CCS for small molecule identification first appeared in the literature in 2010 (Figure 10).[262] In this work, the example molecules, isomeric hydroxylated metabolites of ondansetron, have indistinguishable MS/MS fragmentation spectra due to the hydroxyl moiety being located on the unfragmented benzene ring for all isomers. Thus, the matching of calculated CCS to the experimental CCS distributions was the sole distinguishing dimension of data for identification. Campuzano and colleagues quickly laid the foundation for use of DFT-based CCS calculations in small-molecule identification.[263,264] Recently, a high-performance computing (HPC)-friendly cheminformatics workflow, the *in silico* chemical library engine (ISiCLE), was created to automate all steps of DFT-based CCS calculations and uses software freely available to academics.[265]

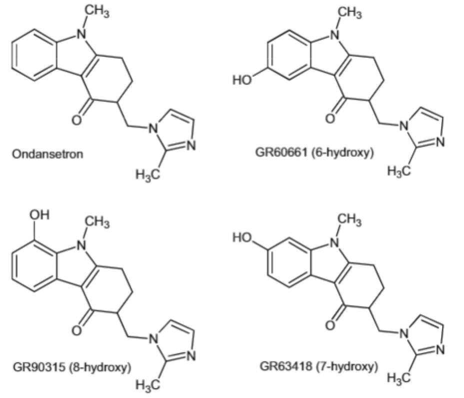


Figure 10. Drug ondansetron and the three hydroxylated metabolites identified in Dear *et al.* 2010.[262] This study was a first demonstration of the use of quantum chemical-based theoretical calculations of CCS for small-molecule identification. The four isomers had indistinguishable MS/MS fragmentation spectra due to the hydroxyl moiety being located on the unfragmented benzene ring, thus, the matching of calculated CCS to the experimental CCS distributions was the sole distinguishing dimension of data for identification. Reproduced with permission from ref 262. Copyright 2010 Wiley and Sons, Ltd.

2.3. Infrared Spectroscopy

IR spectroscopy has a long history of use in the identification of small molecules. During the past 30 years, hundreds of papers have described how IR spectroscopy techniques can be coupled to quantum chemical calculations like DFT to aid identification (via predicted resonance frequencies) and quantitation (via prediction of peak intensities). Like CCS, IR spectra can provide additional evidence for specific molecular structures, which may not be differentiated through traditional MS methods. For example, isobars, enantiomers, and other types of isomers may not always be separated using reversed-phase chromatography and may have identical MS/MS fragmentation spectra[266] or even indistinguishable CCS. With rare exception, IR spectra are unique for every small molecule.[267] Note that IR-vibrational circular dichroism (VCD) methods are required for determining the stereochemistry of chiral molecules. This approach, too, is amenable to DFT calculations that are accurate enough to determine absolute configuration.[268]

Martens *et al.* provided a thorough review of IR ion spectroscopy (IRIS) for small-molecule identification, including discussion of the history of IRIS for molecular identification, IRIS techniques, experimental advances, and applications in untargeted metabolomics.[269] Furthermore, this review highlights the use of DFT, which can be both fast and accurate for gas-phase organic molecule ions, to support identification of unknowns without having a physical reference (*i.e.*, standard) available. As an example, Martens *et al.* demonstrate the identification of a feature unidentifiable by LC-MS/MS, generated from samples taken from two siblings with a neurological disease of unknown etiology. IR spectra were calculated for candidate molecules obtained from a METLIN[270] search, and the feature was identified as methyl-2-pyrrolidinone (an industrial solvent used in the production of polymers) by matching the measured IR spectra to a predicted spectrum (Figure 11).

Of significant importance to the small-molecule identification field is the recent advancement of cryogenic IRIS, spearheaded by groups at the Swiss Federal Institute of

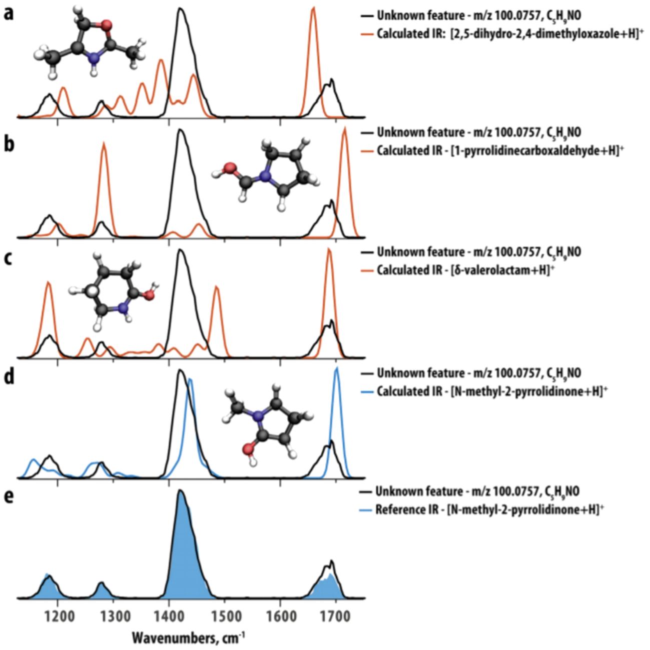


Figure 11. Computed IR spectra (colored traces, a−d) of potential candidate structures resulting from a database search for an unknown feature at *m*/*z* 100.0757 compared to the IRIS spectrum of the unknown feature (black trace) from the patient sample. (e) Compares the IR spectrum of the reference compound *N*-methyl-2-pyrrolidinone identified by the match found for the predicted spectrum. Reproduced with permission from ref 269. Copyright 2020 Elsevier.

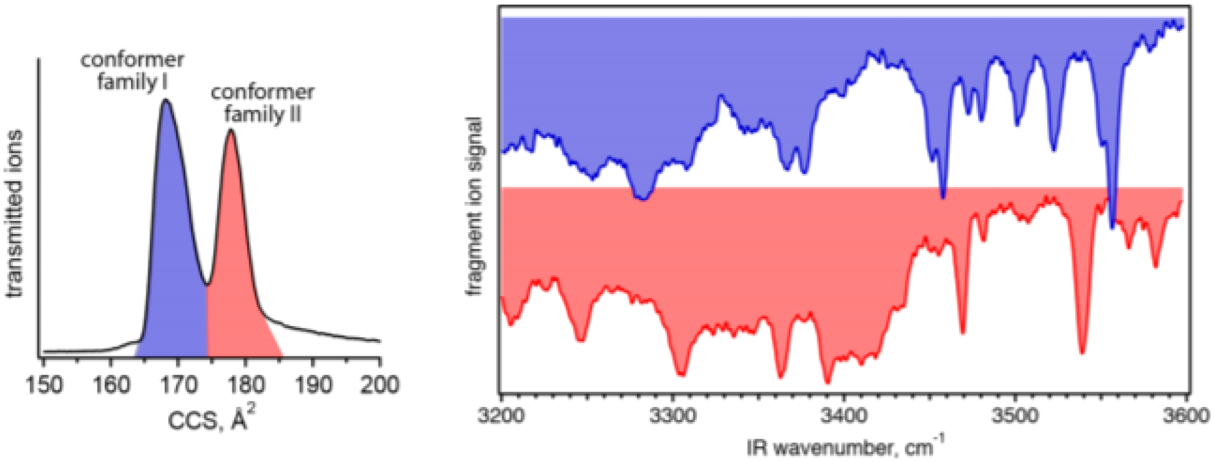


Figure 12. Ion mobility data (left) and cryogenic infrared vibrational spectra of individual conformer families of bradykinin $[M + 2H]^{2+}$, consistent with the *trans*-Pro^2/*trans*-Pro^3 isomer geometry. Reproduced with permission from ref 261. Copyright 2018 American Chemical Society.

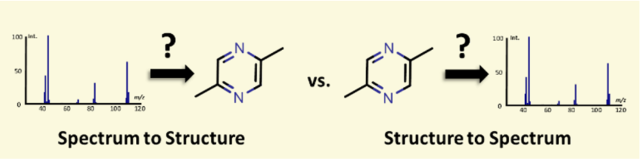


Figure 13. MS-based compound identification paradigm. Mass spectrometrists mostly deal with unknown mass spectra that need structural assignments (spectrum-to-structure). New algorithms to understand fragmentations need to be developed with the help of the quantum chemical community. Since the development of Grimme's QCEIMS method in 2013,[290] it is now possible to predict 70 eV mass spectra using Born−Oppenheimer *ab initio* molecular dynamics directly from structures (structure-to-spectrum). Large chemical databases with millions of compounds can be used to predict high-quality theoretical mass spectra.

Technology Lausanne and University of Florida, led by Professors Thomas Rizzo[260,261,271−276] and Nicolas Polfer,[277−281] respectively. If the ultimate goal of IR spectroscopy is to determine structural information on a molecule, the highest resolution (and highest deconvolution) can be obtained by reducing thermal inhomogeneous broadening and conformational heterogeneity.[260] This can be done by cooling the ion packets (cryogenic cooling) and using conformational selection techniques (such as ion mobility). Cryogenic IRIS is sensitive to small changes in conformation, and when it is combined with a conformational filter such as IMS, both the measured CCS and IR spectra data are bolstered to provide overwhelming evidence for the presence of a molecule,[261] which can be confirmed through corresponding quantum chemical computations without relying on reference material. This is an emerging technology, and it remains to be seen how IMS, cryogenic IR spectroscopy, and quantum chemical calculations can be coupled for molecular identification to be performed in complex samples.

Some initial results on small molecules appear promising. As the only example that couples all three of these techniques to date, in 2018, Kamrath and Rizzo determined that the 178 Å^2 CCS conformer of the N-terminal fragment of bradykinin (RPPGF; $[M + 2H]^{2+}$ adduct) was consistent with the *trans*-Pro^2/*trans*-Pro^3 isomer geometry (Figure 12).[261] This analysis relied on cryogenic IR spectra and DFT optimized molecular geometries. Interestingly, the *trans*-Pro^2/*trans*-Pro^3 isomer geometry is also that which is observed in solution via NMR and implies that the solution structure maintains its geometry during ESI. Other works without coupled ion mobility have also shown the value of quantum chemical calculations for molecular and conformer identification with the use of conformer-selective cryogenic IR spectroscopy.[275,281,282]

2.4. Mass Spectrometry

The main challenge of mass spectrometrists who are attempting to identify unknown compounds is in the interpretation of tandem mass spectra, thus the spectrum-to-structure approach (Figure 13). Quantum chemistry has had less impact on this subfield of MS when compared to spectroscopic methods such as IR, Raman, or NMR spectroscopy.[283] For example, there are no completely quantum-chemistry-based computational pipelines that allow for truly unknown mass spectra to be analyzed, substructure assembled, and final structures proposed. Additionally, more complex approaches such as multiple-stage MS/MS exist (*i.e.*, MS^n) that allow more intricate relationships between precursors and multiple product ions at different stages to be built.[284,285] So far, only cheminformatics, machine learning approaches,[286−288] or hybrid models have been able to perform successful spectrum-to-structure analysis.[289]

The structure-to-spectrum approach is a very promising method for compound identification, because the number of known chemical structures far exceeds the current number of available experimental mass spectra (∼1 million).[291] Molecular databases such as ChemSpider and PubChem provide around 100 million compounds that could be used to calculate theoretical mass spectra based on quantum chemistry. It would then be possible to search unknown experimental spectra against vast libraries of theoretically predicted mass spectra.

In contrast to quantum-chemistry-based calculation of NMR, IR, UV, and Raman spectra, no straightforward procedure exists for quantum-chemistry-based prediction of mass spectra. For example, prediction of IR and Raman vibrational spectra became possible by 1965 using simple FFs[292] and in the late 1970s using *ab initio* calculations. While single fundamental fragmentations can be predicted with the help of quantum chemistry, the cascade of reactions and rearrangements resulting from multiple reaction pathways, and most importantly the m/z peak abundances from complex molecules, have been highly difficult to deduce.

In a major breakthrough and one of the most important discoveries in computational MS, Grimme published the Quantum Chemistry Electron Ionization MS (QCEIMS) program for the first principle calculation of 70 eV mass spectra in 2013.[290] QCEIMS is discussed in more detail in section 2.4.1.

2.4.1. Electron Ionization (EI).

Electron ionization (EI) MS (70 eV) is an established analytical technique and is commonly coupled to GC for analysis of small molecules below 400 Da. Electrons are emitted from a heated filament and focused on gaseous neutral molecules. When the accelerated electrons hit the neutral molecule, radical cations are formed and another electron is ejected. The vibrationally excited carbocations then undergo further bond dissociations and fragmentations on a very fast time scale. The smaller mostly singly charged fragment ions are then accelerated toward a detector and recorded as spectral signals. The ionization efficiency at 70 eV is the highest, and most molecules can be ionized at this energy, allowing for creation of reproducible mass spectra.[293] The power of GC-MS lies in the fact that the instrument industry has subsequently standardized the EI source energies to 70 eV, resulting in the availability of reproducible spectra and available databases to search.[294] Gas chromatography coupled to tandem mass spectrometry (GC-MS/MS) has not reached a breakthrough yet due to the more complex instrumentation and missing MS/MS spectral databases for spectral matching.[295]

Historically, the interpretation of EI-derived spectra depended on statistical rate theory[296−300] and investigation of kinetic processes, especially work based on quasi-equilibrium theory (QET)[301] and Rice−Ramsperger−Kassel−Marcus (RRKM)[302−305] theory, which can be used to predict rate constants. Many of the classical investigations of 70 eV radical cations or anions are limited to single ion species or specific molecules due to the complexity of fragmentation and rearrangement reactions.

The main disadvantage of traditional QET/RRKM approaches is that rate calculations are based on the selection of specific ion transition states and activated complexes on the PES. With increasing atom numbers, the complexity of the reaction space rises exponentially and would require *a priori* knowledge of reaction pathways that are not always available.[306] Methods such as the global reaction route mapping (GRRM) strategy,[307] the AutoMeKin software[308] or the Chemical Trajectory Analyzer (ChemTraYzer) software[309] have been developed to systematically and automatically explore the reaction space.[159]

The QCEIMS approach published by Grimme in 2013 combines Born−Oppenheimer molecular dynamics (BOMD), a type of AIMD, with statistical sampling to predict 70 eV mass spectra.[290] In contrast to other methods, QCEIMS is purely based on physical and chemical principles and can calculate mass spectra from any given molecule. Using a combination of *ab initio* molecular dynamics (AIMD) and stochastic sampling across hundreds of reaction pathways, the correct m/z value of ions and their associated abundances can be predicted. More excitingly, all reaction trajectories are retained and allow for a "look inside" the reaction processes of a mass spectrometer, which then makes it possible to investigate all fragmentations and rearrangements individually. To achieve a balance between efficiency and accuracy, QCEIMS can calculate on various levels of theory, including semiempirical models OM2/OM3,[310] DFTB+,[311] GFNn-xTB,[312] and several DFT methods. The complex relaxation processes from the electronically excited state of the precursor ion are modeled by limiting the reaction on ionic ground-state PES. The impact excess energy is converted to kinetic energy by a heating process, during which the atomic velocities are scaled to a preset impact excess energy value. Such a simple electronic structure can handle the fragmentation reactions and its ability to give a reasonable result is one of the key innovations of QCEIMS.[290]

The QCEIMS software is coupled with several independent software packages such as ORCA,[313,314] TurboMole,[315] MOPAC,[316] MNDO99,[317,318] and DFTB+.[311] Most importantly, the latest independent and therefore stand-alone version of QCEIMS directly implements the GFN-XTB method. This allows for simple installation and practical use of QCEIMS in any research environment with access to HPC. The only required input is a chemical structure. Because the GFNn-xTB[312] methods are parametrized to elements up to $Z = 86$, they are applicable to the most common molecules and therefore provide calculations of 70 eV mass spectra with metalloids such as silicone.[319,320] This is important because the trimethylsilyl group (TMS) is often used during GC-MS derivatization experiments.[321]

One of the advantages of QCEIMS is that reaction pathways are automatically recorded as MD trajectories during the simulation. This allows for comprehensive investigation of the fragmentation mechanism. However, the confirmation of such reactions would require comprehensive investigations because, for any given reaction, a multitude of possibilities exists. In the original paper,[290] Grimme found that most of the primary fragmentations occur within 2−3 ps, while secondary fragmentation reactions take much longer but are important in larger systems. Many well-known reaction pathways in MS are accurately reported by QCEIMS, including α-cleavage,[306] McLafferty rearrangement, retro-Diels−Alder[322] reaction, and CO loss.[290] For molecules with several tautomers, a combination of initial conditions based on Boltzmann population can be used to improve simulation accuracy.[322]

In 2016, Cautereels *et al.* described a different method for the calculation of 70 eV mass spectra using empirical rules for limiting the number of fragmentations along the PES based on DFT calculations.[323] The rules include observations of bond strengths, bond cleavages (that are thermodynamically controlled), and 1,4-rearrangements and McLafferty rearrangements (that are kinetically controlled).[324] The procedure includes conformational sampling and calculation of Boltzmann weights including the calculation of the most stable radical cations. Homolytic and heterolytic fragmentation pathways are calculated under observation of the heuristic rules. Final peak abundances are determined based on a formula that includes the average of energies of the fragmentation pathways and specific fragments. Such an approach could become very useful in the future for detailed investigations of reaction pathways using classical transition state theory.

Similar to the evaluation of machine-learning prediction methods such as CFM-ID,[325] quantum chemical models have to be rigorously tested by comparing theoretical predictions against experimental reference spectra.[326] Similarity match scores and compound rankings should be reported.[294] This can be done with the National Institute of Standards and Technology (NIST) MS Search program and the NIST and MassBank of North America (MoNA) mass spectral databases.[295]

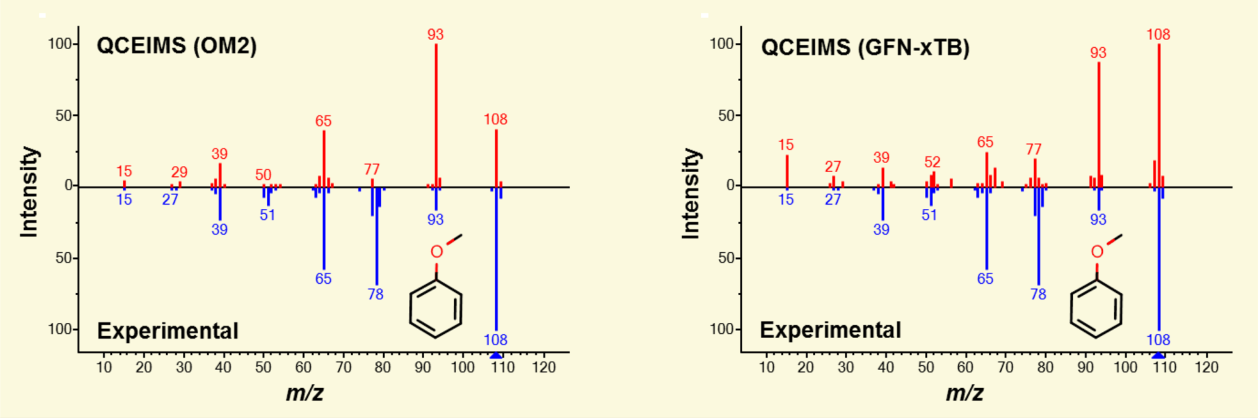


Figure 14. The 70 eV mass spectra of anisole calculated with QCEIMS. (left) The *in silico* spectrum calculated with the OM2 semiempirical function, while (right) shows the GFN1-xTB Hamiltonian. Both algorithms underestimate the peak at *m*/*z* 78 and overestimate the peak at *m*/*z* 93. This leads to a similarity score of 569 for OM2 and a somewhat higher match score of 660 for the new GFN1-xTB method. Further methodic improvements have to be made to increase the quality of the simulated spectra.

2.4.2. QCEIMS Computational Costs and Accuracy.

The QCEIMS protocol contains three types of quantum mechanical calculations: energy/force calculations to generate the potential energy surface for MD, molecular orbital (MO) calculations to determine internal excess energies, and ionization potential (IP) calculations of each fragment to generate the statistical charges. The original version of QCEIMS utilizes DFT methods for MO and IP calculations, whereas the energy/force calculations for the time-consuming MD steps use the OM2/OM3[327] orthogonal corrected semiempirical methods.

For example, the simulation of the 70 eV EI mass spectrum of anisole (C_7H_8O, MW = 108.057 Da) (Figure 14) requires 1.2 million individual MD steps and 82 min of computational time on 16 CPU cores at the OM2 level. The choice of the underlying method significantly affects simulation speed. The GFNn-xTB methods[319,320] will be 10−20 times slower than the semiempirical OM2[310] simulations, while purely DFT-based MD can be 100 or more times slower than the semiempirical methods.

The computational cost for semiempirical methods is usually much smaller than *ab initio* methods (OM2/PM6[328] < GFNn-xTB < DFT), whereas in terms of accuracy, we see the opposite trend with DFT being the most accurate method (DFT > GFNn-xTB > OMx/PMx). Interestingly, chemical bonds are more easily dissociated in semiempirical simulations relative to the more accurate DFT simulations.[329] Therefore, more accurate calculations of the PES may require even more simulation steps with longer fragmentation processes. While semiempirical methods like OM2/OM3 are significantly faster than GFNn-xTB, the PES along bond dissociation coordinates are not sufficiently accurate, leading to simulated spectra that have lower similarity scores when compared to experimental reference spectra (Figure 14).

Because DFT methods are closer to the "exact" PES, they should be used as a reference in evaluating more approximate models,[329] but their increased computational cost puts them out of reach for simulating EI mass spectra of larger molecules. We are optimistic that GPU-accelerated implementations of DFT methods in software such as TeraChem[87,330] or Fermions ++ may lead to fast high-accuracy simulations.[331]

However, the usage of fractional occupation number weighted densities[332] can reproduce some properties of multireference wave functions, making it a possible low-cost alternative for treating multireference systems. On the other hand, when the energy gap between the excited state and ground state goes to zero, the Born−Oppenheimer approximation and single reference methods used in QCEIMS can break down. The treatment of highly excited electronic states using multireference methods,[96−98] such as the states accessed during QCEIMS, is under active investigation and can guide the development of improved simulation approaches in the future.

The accuracy of predicted *in silico* spectra has to be evaluated against diverse and large number of experimentally measured spectra.[294] QCEIMS (with OM2/OM3) performs on the same accuracy level as the best available machine learning algorithms such as CFM-ID.[325] The QCEIMS method also has the advantage that any given molecule can be calculated. The reason is that machine learning methods require experimental training data, while QCEIMS as an *ab initio* method is only based on physical and chemical principles. The most important question for practitioners is the practical use of algorithms in daily research applications. Currently, it is not possible to calculate most compounds with high similarity match scores (>850). It is also not yet possible to determine the quality of predictions in advance due to the stochastic nature of the computations. It is foreseeable that with improved accuracy of future versions of QCEIMS and related methods, a wide range of *in silico* spectra can be used for training in machine leaning to allow for even faster simulation of *in silico* mass spectra from all known compounds.

2.4.3. Coupling EI to Other Spectroscopic Methods.

While GC-MS mass spectra at 70 eV can give structural insights, it is not possible to fully interpret all mass spectra because in many cases the molecular ion is not observed and following individual fragmentations is not directly possible. Techniques such as chemical ionization or cold EI can help increase the stability and abundance of the molecular ion.[333] Furthermore, integrating parallel analysis techniques such as IR, Raman, and UV will allow for easier structure-to-spectrum identification using quantum mechanical calculations of optical spectra.[334] In such a case, MS and optical spectroscopy experiments are performed in parallel, and the resulting spectra can be investigated theoretically using quantum chemistry methods or QM/MM.[335] For example, threshold photo-

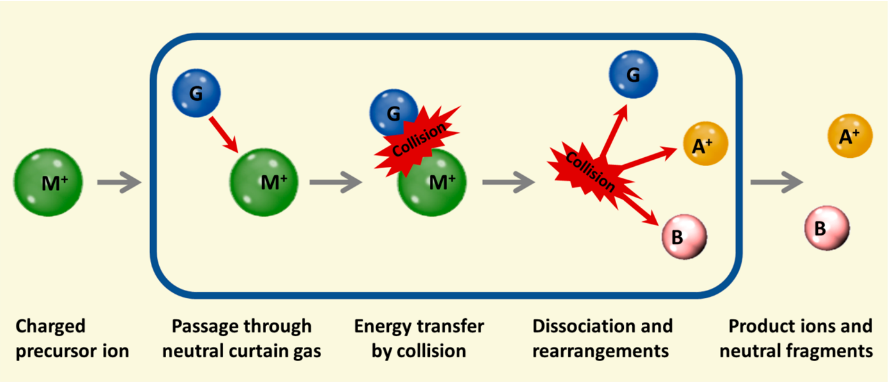

M+
G
M+
G
Collision
M+
G
Collision
A+
B
A+
B
Charged precursor ion
Passage through neutral curtain gas
Energy transfer by collision
Dissociation and rearrangements
Product ions and neutral fragments

Figure 15. Collision-induced dissociation process (CID-MS/MS) in a collision cell. Ions formed by, *e.g.*, electrospray ionization enter the mass spectrometer and pass a neutral curtain gas (He, Ne, Ar). Once energy is transferred from the collision, molecules can dissociate or rearrange. Ions are further accelerated toward the detectors and registered as specific *m*/*z* signals. The collision gas and neutral reaction products are removed by the vacuum pumps of the instrument.

ionization mass spectra can be acquired with photoelectron photoion coincidence (PEPICO) spectroscopy and can be coupled with DFT calculations to gain insights into fragmentation behavior.[336−338] In particular, coupling MS with IR multiple-photon dissociation spectroscopy (IRMPD) seems to be an excellent way for interpreting dissociation pathways by combining experiments with quantum chemical calculations.[337] While such instrumental setups are complex and expensive, they show the possibilities of instrumental integration with quantum mechanical computations. Such techniques, while discussed here in detail for EI, can also be coupled to other methods such as ESI and CID MS/MS.

2.4.4. Collision-Induced Dissociation.

Tandem mass spectrometry (MS/MS) uses one or more mass analyzers paired with a fragmentation technique that activates and dissociates ions to identify structural information. The tandem-in-time concept describes fragmentation processes in ion trap and Fourier transform ion cyclotron resonance (FT-ICR) mass spectrometry instrumentation, whereas the tandem-in-space describes fragmentations in triple-quadrupole (QQQ), quadrupole-time-of-flight (Q-TOF), and hybrid instruments.[339] The MS/MS approach is commonly applied to compound identification in mixtures. In this procedure, a soft ionization method such as ESI is generally used in the first stage to generate nearly intact molecular ions from the sample, which are selected by mass in the first mass analyzer. These ions are then fragmented via ion activation processes, followed by subsequent mass analysis to determine their structural composition. While many other ion activation modes such as surface-induced dissociation or in-source fragmentation exist,[330] only the most prominent fragmentation method, CID, also sometimes called collision-activation decomposition (CAD), is discussed here.

CID was introduced by the Jennings and McLafferty groups in 1968 as a tool to identify the structures of molecular ions.[340,341] During the 1970s, CID-MS/MS was made commercially available in a form of sector-MS/MS, and later triple quadrupole mass spectrometers were developed by Yost and Enke.[342,343] Initially, most experiments were performed using high collision energies in the keV range, but today, instrumentation operating in that regime is only used for very specific research purposes.

In the 1980s, lower energy CID-MS/MS with collision energies in the 1−100 eV range was introduced by Douglas and Dawson.[344−346] The majority of today's commercial mass analyzers operate in the low energy CID mode. The most common instrumental setups for untargeted profiling of complex biological samples include ultraperformance LC coupled to orbital ion trap tandem mass spectrometers (Orbitrap) or quadrupole-time-of-flight instruments (Q-TOF). Compound identification for small molecules mostly relies on database search of low-energy CID-MS/MS spectra[295] or HCD (higher-energy collision dissociation)-MS/MS spectra from orbital ion traps.

Currently available MS/MS databases are rather small, with fewer than 500 000 covered compounds, and metabolic profiling reports still contain many unknown MS/MS spectra.[347] Therefore, a major research topic is the modeling of high-quality CID-MS/MS *in silico* mass spectra directly from compound structures (*structure-to-spectrum* approach). The spectral interpretations can provide a better understanding of the CID fragmentation pathways in unknown CID-MS/MS spectra (*spectrum-to-structure* approach).

The CID process can be distinguished into different types: collision-cell CID, ion trap-CID, and in-source CID.[348] During the collision-cell CID process, ions pass through the collision cell, a vacuum chamber that is flooded with inert gas such as helium, nitrogen, or argon, where binary collisions occur in the gas phase (Figure 15). The collisions within the cell increase the internal energy of the ions and enhance the rate of the unimolecular fragmentation reactions that create the product ions. If the amount of energy transfer is great enough to break chemical bonds, the ion consequently decomposes into charged ions and neutral fragments. In the case of collision-cell CID, an ion makes approximately 10 collisions on average while traveling through the collision cell within 20 μs; as many as 100 collisions are possible in ion-trap CID over the corresponding residence time of ~5 ms.[349,350]

Although the entire CID process occurs on time-scales that are inaccessible by *ab initio* or even semiempirical MD methods, it may be possible to use simplified models to approximate the kinetic energy that is imparted by non-dissociative collisions that lead up to the dissociative event, in a similar way to how QCEIMS uses a velocity rescaling to model the complex energy conversion process from an electronic excited state to nuclear kinetic energy. Like the case of EI-MS, there are many possible pathways for collisional activation and fragmentations across the diverse configurations of CID experiments and instrument configurations. Many of them should be considered during quantum chemical investigations.

2.4.5. Interpretation of CID-MS/MS.

The CID-MS/MS process includes many different fragmentation and rearrangement reactions, divided into charge retention and charge migration mechanisms.[351,352] For quantum-chemistry-based CID-MS/MS investigations, one can distinguish between statistical and dynamical models. Specifically, (i) transition state theory and simulations along the PES and (ii) molecular dynamics (QM/MM) simulations.[353−356]

Semiempirical and DFT models have been used in the past on a number of molecular species to understand CID-MS/MS dissociations.[357−359] For example, the number of bond cleavages and proton migration events were determined by semiempirical methods.[360] The QC-FPT method was used to predict thermodynamically feasible fragmentations across a number of sample molecules.[361]

The VENUS software, developed by the Hase group at Texas Tech University, has been extensively used to model CID MD processes. VENUS has been interfaced with software tools for semiempirical (MOPAC)[362] and DFT calculations (NWChem).[363] Several papers also have shown that MD simulations can be coupled to the automated reaction mechanisms and kinetics software AutoMeKin/TSSCDS, which automatically explores transition states.[308,364,365] Because of the time-consuming nature of the approach, mostly single molecules such as sugars, sterols, amino acids, and peptides have been described.[365−371]

A QM/MM direct dynamics simulation described in a recent paper further details the modeling of center-of-mass collisions of argon and galactose-6-sulfate.[372] The reaction process can be visually investigated, including collisions and formation of multiple fragments. This computational process with improved statistical sampling can be considered as a foundation for simulations of CID-MS/MS spectra. Further details of the historical developments and implementations of QM/MM methods are comprehensively described in the excellent book by Song and Spezia.[373]

2.4.6. Prediction of *in Silico* CID MS/MS Spectra.

As of 2020, no commercially or publicly available quantum chemistry method exists with the ability to *accurately* predict CID-MS/MS spectra from *diverse* compounds (<1000 Da). While several heuristic, machine learning, and reaction-based models can accurately model CID-MS/MS spectra for small molecules, they usually require experimental training spectra and are discussed elsewhere.[288]

Similar to CID fragmentation predictions, classical reaction mechanism theory as well as QM/MM calculations including AIMD can be used to predict ions and their abundances. The QCMS^2 approach uses a pipeline of conformational sampling, protonation, fragmentation rules, and exploration of transition states along the PES.[374,375] Formerly developed for EI-MS, it has been modified to work for CID predictions.[376]

A QM/MM chemical dynamics approach, using the VENUS chemical dynamics software and semiempirical and DFT calculations, was able to predict fragments and relative peak abundances of testosterone.[366] Another workflow used an automatic tautomerization network and was subsequently able to describe a mobile proton model during the fragmentation of 6-*O*-methylguanine.[377] Fragments and relative abundances were calculated and compared to experimental Orbitrap HCD-MS/MS spectra. Several efforts have been made to predict CID-MS/MS spectra for specific target compounds, mostly with constraints on compound diversity or the experimental setup.[373]

While fragmentations may be easily predicted,[357,358,378] the prediction of ions with a low false-positive rate and accurate prediction of associated ion abundances are still very challenging.[366] In addition, many other aspects of automatic CID-MS/MS modeling workflow need to be considered, such as correct sampling of protomer, rotamer, and conformer samplings, as well as energy distribution and transfer upon collision with various collisional energy.[361,379] Some of these challenges are further discussed below.

Adduct formationis of high interest for mass spectrometrists because a multitude of adduct ions such as $[M + Na]^+$, $[M - H]^-$, or $[M + Cl]^-$ are observed during experiments.[380,381] Depending on the ionization mode, solvent, and buffer systems, different ions are formed during the electrospray process[295,382,383] that can strongly influence the MS/MS fragmentation process and observed product.[295,384,385] A statistical analysis of the NIST database on 80 000 MS/MS spectra covering 300 possible adduct species showed that protonation $[M + H]^+$ and deprotonation $[M - H]^-$ are the most commonly observed adducts in ESI. Doubly or higher charged species are rarely observed in small molecule MS/MS spectra, but they are very important in proteomics and peptide sequencing.[386,387]

Multiple protomer species must be considered in the cases of complex molecules with diverse ionizable groups.[388−390] The selected computational approach must be able to model intramolecular proton migrations.[391] While the best practice currently is to calculate protomer ensembles using a Boltzmann distribution,[357] experimental investigations have shown that the observed protomer species might not always be the thermodynamically favored species.[392] The freely available Conformer−Rotamer Ensemble Sampling Tool (CREST), a conformer-rotamer ensemble sampling tool based on the GFN-xTB method, can automatically enumerate and compute energies of multiple protonation states.[379,393] The included ensemble sorting tool (CREGEN) then performs a Boltzmann population analysis and ensemble analysis. While all of these computations are based on the semiempirical GFN-xTB level.[394] Orthogonal techniques such as ion mobility spectrometry (IMS) can help to understand multiple protonation sites.[395]

The modeling of tautomers can also be challenging.[396] While tautomers have been observed in many mass spectrometric experiments,[397−400] the formation process is solvent- and pH-dependent and not fully understood yet.[377,401,402] However, tautomers can be easily enumerated with commercial and open cheminformatics tools such as ChemAxon, OpenEye, RDKit, MolTPC, CDK, or AMBIT.[403−406] A fully automatic generation of tautomers using solvation models and quantum chemistry approaches is possible with CREST and the XTB software.[394,402]

The influence of conformer ensembles, rotamers, and stereoisomers is important for predictions across many spectroscopic methods. However, MS alone is rarely used as standalone technology to analyze rotamers or stereoisomers because additional orthogonal separation technologies such as GC, LC, or ion mobility allow for better resolution.[407] The consideration of multiple conformers is important for CID modeling. The existing software for generation of conformer ensembles for NMR spectral predictions could be used and integrated into CID modeling workflows.[210,408] The freely available XTB, CREST, and ENSO software modules are an excellent starting point to perform such analysis.[393]

The difficulties of modeling CID-MS/MS spectra are based on a number of factors, including diverse instrumental setups,[409] different ion activation schemas, and CID-voltage dependent fragmentations.[410] Also, the time-scale of fragmentations are important. Whereas CID collisions are modeled in the picosecond to femtosecond range, the experimental time-scale can last up to milliseconds in certain instruments such as FT-ICR MS.[378,411] Modeling a single collision might not be sufficient because multiple collision events may occur depending on instrument type.[411,412]

The developmental focus for the structure-to-spectrum approach should be on modeling low-energy CID-MS/MS spectra that match the most commonly used instrument types. That includes QTOF-type (CID-MS/MS) and Orbitrap-type (HCD-MS/MS) instrumentation for the small-molecule communities.[413,414] Modeled *in silico* spectra must contain accurate m/z values and associated peak abundances. The quality of prediction needs to be reported by matching theoretical MS/MS spectra against experimental reference spectra visualized in head-to-tail view, using cosine, dot-product, or other mass spectral match scores.[295] There are many public and commercial MS/MS databases, including MoNA, NIST, METLIN, and mzCloud, that can be used to obtain experimental reference MS/MS spectra.[291] The publication of *opaque* approaches should be avoided, describing methods or software that are neither commercially nor freely available. *In silico* spectra should be made publicly available in electronic (MSP/MGF) format.[295,415]

Overall, methods for accurate prediction of CID-MS/MS spectra by quantum chemistry approaches are still in their infancy. However, current advances for *de novo* prediction of MS/MS spectra made by multiple groups worldwide are very promising, and the topic itself is currently gaining attention within the quantum chemistry community. Looking even further into the future, quantum computers will be able to solve quantum chemistry problems[416] and predict molecular spectra[417] at an unprecedented speed and scale. Theoretical frameworks must be developed and linked to perform confident spectrum-to-structure and structure-to-spectrum predictions in the realm of MS to take advantage of this advance.

3. CURRENT BEST PRACTICES AND KNOWN SOURCES OF ERROR

3.1. Basic Procedures and Implementation Methods

3.1.1. Containerization and Cloud Computing. Traditionally, quantum chemists use local HPC resources, including compute clusters and local data storage. Considering cloud resources, one should carefully consider a number of aspects: computing costs, data storage, and data transfer, such as egress costs (moving data out of the cloud).[418] This could result in substantial charges in cloud computing that have to be covered by a research budget, whereas local computing is mostly provided for free or is covered by low-cost contributions. The NIH Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative is actively investigating cloud storage and cloud computing with two major partners, Google Cloud and Amazon Web Services (AWS). The aim is to use cost-. and industry-leading commercial resources to advance biomedical research (datascience.nih.gov/strides).

There are many advantages to using commercial cloud services. Microservices and serverless computations (AWS lambda) and easy containerization using Docker or Singularity allow for quick deployments.[419,420] Furthermore, horizontal scaling to thousands of instances is easily possible on the cloud, while local HPC clusters usually have limitations in available on-demand compute capabilities. Customizing CPU and memory use and the use of preemptible instances can greatly reduce the costs.

While local HPC centers have administrative staff and programmers, when using cloud services, the user in many cases also becomes the systems administrator. This can be initially problematic when setting up scheduling systems such as Torque PBS or Slurm or the need for setting up complex networks or cloud data-storage solutions. Finally, licensing requirements must be cleared with commercial software providers when hundreds or thousands of instances are deployed or software is limited in terms of CPU sockets in use. Here the use of open-source quantum chemistry software, such as Psi4[82] and NWChem,[84] can be recommended with no distribution or licensing restrictions.

3.2. Sources of Error and Methods of Correction

The models we use to represent the world are not perfect; however, quantum chemistry has come a long way in being able to accurately reproduce many aspects of the natural world. Even so, errors persist. Below, major sources of errors computational chemists face when performing quantum mechanical calculations are addressed. For additional in-depth discussions on errors in theoretical calculations on small molecules, see the following references (135,231,421−423).

3.2.1. Level of Theory. In general, the accuracy of quantum mechanical calculations improves as the cost (measured, generally, in terms of computational time) of the calculations increases. To obtain accurate results without excessive computational cost, DFT is generally recommended, as opposed to less accurate but faster levels of theory, such as HF,[424,425] or generally more accurate but time-consuming levels of theory, such as CCSD.[135,231,421,422,426−431] Goerigk and co-workers, in both their 2017 and 2019 reviews of "the density functional theory zoo," stress the importance of including dispersion corrections, such as D3(BJ) and suggest using double-hybrid functionals.[135,421] However, depending on the specific research questions to be addressed and the acceptable error bars that can be tolerated, the well-worn and too often maligned B3LYP method is often sufficient.[231,427,432−434]

3.2.1.1. Solvent Effects. Solvation can be represented both implicitly (with a field) and explicitly (with discrete solvent molecules included in the calculations). Implicit solvation is a good starting point because it is much less computationally demanding than explicit solvation modeling and is frequently sufficient to calculate desired properties. Explicit solvation should always be considered, however, if strong intermolecular interactions are involved, *e.g.*, traditional H-bonding, CH−X, cation−π *etc.*[231,435] For example, Da Silva *et al.* found including four explicit chloroform molecules along with an implicit chloroform model afforded more accurate ^{1}H chemical shifts than with the implicit solvent model alone.[435] This is but one representative example.

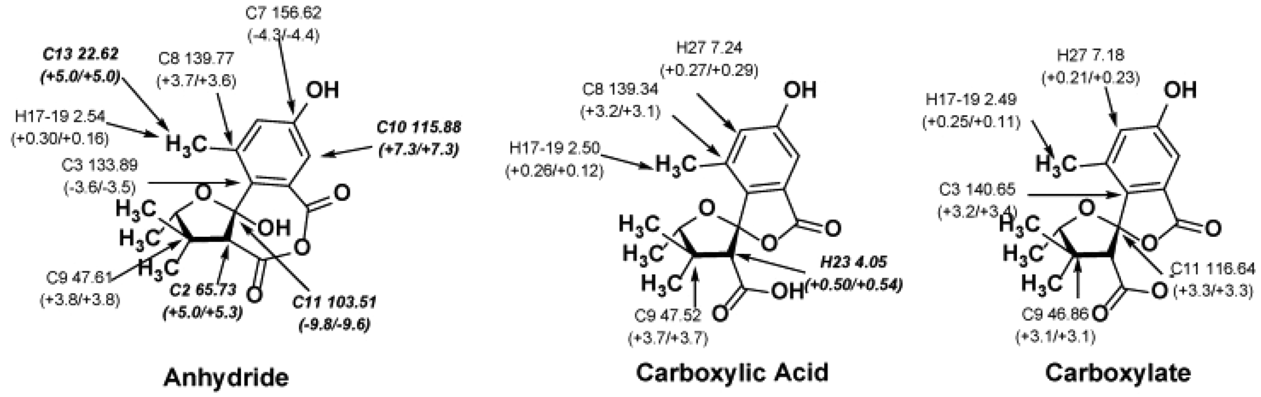


Figure 16. Effects of a labile carboxylic acid proton on ^{1}H and ^{13}C chemicals shifts.[436] Reproduced with permission from MDPI (2017, CC-BY 4.0 license).

3.2.1.2. Questions of Dimerization, Acidity/Basicity, Heavy-Atom Effects, and Hybridization. The best approach to modeling the solute is often nontrivial, for example: (i) if a compound of interest exists as a dimer in solution and must be modeled as such to reduce error,[437] (ii) if the protonation state of a compound is difficult to predict and multiple protonation states must be modeled,[436] (iii) if the effects of intramolecular noncovalent interactions on a particular property are overestimated in the calculations an error that potentially can be reduced through randomly averaging out free energies of the conformers as recommended recently by Sarotti and coworkers,[438] or (iv) if the properties of individual atoms, such as labile protons (Figure 16) or atoms next to heavy atoms, which suffer from relativistic effects, require specialized methods.[231,439] For example, Kutateladze and Reddy were able to reduce this latter error by accounting for spin−orbit effects by developing the DU8+ method to predict accurate ^{13}C chemical shifts for ^{13}C−X carbons. With this technique, they were able to correct the structures of numerous natural products with the misassigned configurations or halogen positions, as in the case of tristichone C (Figure 17).[439] Another method used to account for errors specific to atom types is the multistandard approach that has been used to reference different ^{13}C or ^{1}H chemical shifts based on the hybridization of the carbon atom.[440,441] Overall, the more accurately a chemist can model what is occurring in the physical sample that is measured, the more accurately they will be able to predict the chemical properties of their system. Determining what is occurring in a sample from the empirical data, however, generally relies on the chemical intuition of the analyst.

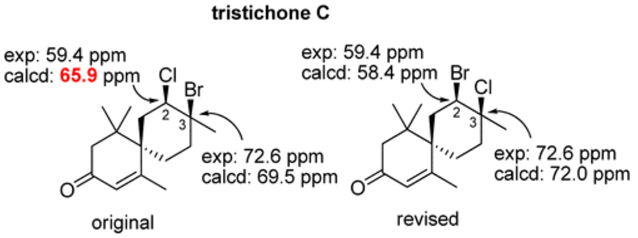


Figure 17. Structural reassignment of the natural product, tristichone C by Kutateladze and Reddy using their DU8+ method.[439] Reprinted with permission from ref 439. Copyright 2017 American Chemical Society.

3.2.1.3. Conformational Searching. Conformational space increases rapidly with degrees of freedom in small molecules, and as such, determining all relevant conformers becomes increasingly difficult as the number of atoms increases.[434] Often, errors are particular to the software and/or conformer generation algorithms used. Traditional approaches, including the use of simulated annealing[442,443] or brute force sampling methods,[444] can introduce errors because these are generally carried out using nonquantum mechanics methods.[445] For example, when searching for conformers of strained rings with some methods, only conformers with the same dihedral angle sign as the original input structure will be found in the conformational search.[446] A new conformational searching method created by Grimme *et al.*, CREST, provides a fast alternative method that addresses this and other errors found in traditional conformational searching methods.[210,445,447] This method, while it does not always provide the most accurate conformer population, often provides adequate sampling of the conformational space and is freely available on Github.

3.2.1.4. Energy Calculations. Calculating accurate relative free energies of unique conformations is important to correctly predict most properties of small molecules because small differences in free energy correspond to large differences in observed conformational populations ($\Delta G = -RT \ln K$)[438,448] and, by extension, the property of interest. Although DFT is often used to calculate a variety of chemical properties, Bootsma and Wheeler recently brought to light the problem that the orientation of an initial molecular structure for an optimization calculation can drastically affect the resulting free energy of that compound if the integration grid (*i.e.*, the number of points sampled) used is too small. With a large enough grid size (*e.g.*, 99 590 or greater), this error can be reduced to 1 kcal/mol or less,[449,450] so using such a grid is suggested for quantum chemical calculations in general.[407,408,427,449,450]

3.2.1.5. Boltzmann Averaging, Conformational Weighting, Conformational Analysis. When performing conformational analysis, depending on the chemical system and properties of interest, a single low energy conformer may be all that is required to arrive at accurate predictions. For more flexible systems, Boltzmann averaging is commonly used to combine per-conformer properties based on the relative populations of these conformers. This approach involves giving a weight to the properties for each conformer with relative free energies within ~3 kcal/mol of the lowest energy conformer (for room temperature calculations).[231] However, the errors in the free energy calculations for each conformer, which are affected by the level of theory, temperature, and solvent model used, can be large enough to significantly affect properties predicted based on this weighting.[135,238,451,452] Consequently, alternative methods are becoming increasingly employed, such as the Computer Aided Structure Elucidation-3D (CASE-3D) method or the random ensemble method.[438,448]

3.2.2. Sources of Error Specific to Techniques.

3.2.2.1. NMR. Computational prediction of NMR chemical shifts and coupling constants has become an invaluable tool for the structure elucidation of natural products.[231,432−435,453−456]

Linear Regression: Simply referencing chemical shifts to a single standard (*e.g.,* TMS for organic, nonpolar, solvated samples; sodium trimethylsilylpropanesulfonate, *i.e.,* DSS, for water or polar solvated samples) can introduce inaccuracies, so linear scaling factors are often used instead to convert isotropic shielding constants to chemical shifts.[231] Scaling factors for ^{1}H and ^{13}C,[231,235,238,457] ^{11}B,[458] ^{15}N,[458] ^{19}F,[459] ^{31}P,[460] ^{13}C−X,[439] and ^{1}H−^{1}H *J*-couplings[461] are available to reduce systematic error. Many scaling factors and instructions on how to use them are freely available on the CHESHIRE (CHEmical SHIft REpository) Web site (cheshirenmr.info).[231] While linear regression allows for accurate scaling in most cases, the multistandard approach or the use of secondary linear regression lines can be used to improve accuracy for specific types of atoms.[238,440,441]

3.2.2.2. pH Effects on Chemical Shifts. pH of the NMR solution also can affect chemical shifts.[435,462,463] How best to model a compound when this is an issue is unclear, although improved predictions have been made by explicitly modeling solvent or acid molecules.[435,462,463]

3.2.2.3. IR/Raman. In IR calculations, anharmonicity can introduce a significant source of error depending on how it is calculated (if at all). The generalized second-order vibrational perturbation theory (GVPT2) is recommended and often used to reduce the effects of Fermi resonances resulting from anharmonicity if highly accurate IR spectra are needed.[427,437]

3.2.2.4. CD. Circular dichroism is again dependent on correct conformational analysis.[434,452,464] In addition, as for NMR, a scaling approach can be applied, here using a wavelength correction.[434]

3.2.3. Limitations and Challenges. As far as quantum mechanical calculations have come in recent years for structure elucidation and small molecule identification, they are limited by the quality of the experimental data obtained for metabolomics studies. Thus, a movement has been made toward including raw data and error bars in publications to allow for better analysis and comparison of scientific data.[423,465] Even with excellent experimental data, the computational chemist is limited by time and resources. With the technology currently available, cyclopeptides with five or so residues represent a practical size limit for accurate DFT-based computations.[466−470]

Other than size, strong inter- or intramolecular interactions, both charged and uncharged, can create challenges in producing accurate theoretical results. Even the seemingly simple question of protonation/deprotonation is often complicated to model appropriately.[436] Because of these factors and the time involved in DFT calculations, automated structure elucidation with NMR spectroscopy is gaining popularity and can be a good starting point for metabolite identification, although it is recommended to confirm the results with quantum chemical calculations. Some popular methods include CASE,[471−473] WebCocon,[474,475] and Logic for Structure Determination (LSD).[476]

3.3. Validation Methods

3.3.1. Statistical Verification with Large Data Sets. Statistical verification for quantum chemical calculations is becoming increasingly popular for DFT calculations of energy. While optimizations performed at efficient levels of theory can provide accurate conformational minima, they often have difficulties accurately predicting relative free energies. Including a statistical component, as with the DP4/DP4+ and related methods, allows one to assess the likelihood of a specific structural assignment being the correct one out of a group of defined possible structures;[228,467] however, it should be noted the DP4+ method does not always suggest the correct stereochemical assignment.[456]

A recently published method, the *In Silico* Chemical Library Engine (ISiCLE),[265] has shown promise in structure determination using NMR, IR, and CCS predictions and has performed well in large-scale compound identification tests, such as the ENTACT interlaboratory challenge.[248] This method, which can be used as a cross-validation method, aims to provide a means for structure prediction without standards, a major advantage to rapidly accelerating compound identification.[265]

3.3.2. Cross Validation. Cross validation with multiple methods can help one avoid errors. For example, comparing results from Boltzmann averaging with Sarotti's recently introduced random ensemble method can reduce errors.[438] Scaling factors are generally cross-validated with a second set of compounds after they are calculated from a test set.[231] In the long run, "consensus scoring," which is frequently done in fields like automated ligand docking, may have an important role to play in compound identification.

4. OUTLOOK FOR ROLE IN METABOLOMICS

4.1. Advancing Standards-Free Metabolomics

Most metabolomics studies use databases with known experimental MS or NMR data obtained from analyses of standard compounds. Such databases are useful but do not go beyond known metabolites. The problem is that a very large number of metabolites are not represented in databases. Identification of such unknowns requires considerable experimental effort using approaches employed in natural products chemistry.[60] Furthermore, the number of purified or synthesized chemicals that are available for traditional database construction will always be a limitation. We believe that computational methods, most specifically those based in quantum mechanical calculations (and eventually coupled to deep learning and quantum computing approaches), can provide considerable help in the identification of unknown compounds.

Using high-level calculations like those outlined in this perspective, QM calculations could be accurate substitutes for experimental data. In fact, QM is a hallmark for high confidence spectroscopic (and spectrometric) simulations, and it has been proved to produce reliable results enabling structure confirmation,[477,478] distinction between isomers,[479] and to point out wrong assignments from peer-reviewed publications.[480,481] One advantage of using QM calculations for database entries is that they can be applied to several experimentally accessible measurements including retention time, MS/MS fragmentation, collisional cross section, and NMR data. One rapidly growing application area that has yet to incorporate standards-free and QM calculation-based identification methods is metabolomics imaging. MS imaging, specifically, has historically utilized m/z features alone for putative identifications, but as MS/MS and IMS are increasingly incorporated into these analyses,[482−487] *in silico* libraries may help to accelerate the usefulness and impact of metabolomics imaging research by providing high confident metabolite identifications.

For NMR data, most experimental databases rely on chemical shift values, but other experimental measurements are highly informative, especially J coupling constants. Both chemical shifts and J couplings can be computed and could be used in matching experimental data. Several types of NMR experiments can be used to measure homo- and heteronuclear scalar coupling constants through one or more bonds,[488] and if the data could be effectively used, then it could be more routinely measured in a metabolomics workflow. Currently, unknown identification of molecules perhaps relies *too much* on the chemical shift alone. This is especially problematic given that ^{1}H chemical shifts can be dependent on several variables including solvent, pH, ions, and temperature. Other nuclei, including ^{13}C,[489] are less dependent on these variables. A combination of computed ^{1}H and ^{13}C chemical shifts, along with J coupling values, would be very valuable for increasing our confidence in unknown metabolite identification. This type of information is now available through QM calculations coupled to spin dynamics simulations.[490,491]

While it would be useful to have a large database of results from high-level QM calculations of metabolites and other small molecules, that is still impossible because of the time required for each calculation, especially for flexible molecules that require conformational averaging. Prioritizing which compounds to calculate is a challenging question. The Brüschweiler lab has developed an interesting approach called Structure of Unknown Metabolomic Mixture components by MS/NMR (SUMMIT) that combines high-resolution MS and NMR data with computation.[492] In SUMMIT, high-resolution MS data are used to determine the molecular formula of an unknown peak of interest. Next, a database of known structures such as Chemspider[493] is used to generate all possible molecules consistent with the experimental molecular formula. Then, NMR chemical shifts are computed for each potential molecule, and computational results are compared with experimental NMR data of the same sample. The primary difficulty with the SUMMIT approach is the large number of possible molecules as the molecular weight increases. There are at least two approaches to deal with this problem. First, one could use lower-level theory that could be effectively applied to hundreds or thousands of candidates. However, as the level of theory decreases, so does the accuracy of the calculation, so matching the correct one against experimental data would be difficult. Another idea is to add an additional layer of filtering, such as additional chemical data such as CCS or biological/genetic knowledge of pathways. if genomic data can be incorporated and related to the unknown feature of interest.

QM simulations of spectroscopic data of organic compounds is ripe to be used for compound identification within metabolomics. Of course, this application will only be really accomplished if a standardized protocol is adopted and an open-access computational database is established. Furthermore, experimental advances are required to make cryogenic gas-phase spectroscopy commercially available, which would offer an ideal pairing with QM calculations. Foreseeing the future advancement of quantum computers, one can easily expect a huge development in high-level spectroscopic simulations. Computing time will be reduced in orders of magnitude and an increasing number of variables (such as solvent, temperature, pH, matrix effects, *etc.*) could be implemented within the calculations. Thus, to have a seed already in place for a community-driven effort to catalogue simulated data in a common database will be a shortcut for the success of many other life-sciences-based studies.

4.2. Comparison to Machine and Deep Learning

Recent advances in deep learning[494−496] have shown promise in predicting physiochemical properties previously reliant on quantum chemical calculations, such as CCS, NMR chemical shifts, and MS fragmentation patterns,[209,497−500] as well as replacing quantum chemical calculations entirely.[501−504] Deep learning has demonstrated improvements in property prediction accuracy compared to quantum-chemistry-based approaches, while reducing per-structure computation time by orders of magnitude (hours for quantum chemical, *versus* milliseconds with deep learning after training).[497,505] Similar improvements are seen with deep-learning-based DFT: orders of magnitude reduction in computation time with sub-1% MAE.[502] For generative approaches, including autoencoder and adversarial networks,[505−516] deep learning offers additional potential in addressing the inverse quantitative structure−property relationship (QSPR) problem,[517−519] wherein molecular structure candidates can be determined from physiochemical property/properties. However, with deep learning follows considerations to training time, which better contextualizes per-structure computation time, and generalizability, or to what extent does prediction accuracy degrade as inputs vary from the training set.

4.2.2. Generalizability. The greatest concern when evaluating deep learning models against first-principles approaches is generalizability, the model's ability to operate successfully outside of the data set on which it was trained. With first-principles approaches, calculations are based on the underlying physics of the quantum-chemical system and are thus less biased to the structures seen, or not seen, during training. As such, a completely novel structure, when simulated by a first-principles approach, is beholden to low-level physical constraints/laws. A spurious prediction is unlikely, as the underlying physics remains constant across chemical space, both known and, presumably, unknown. With a poorly generalized deep learning model, an evaluated structure that exists outside the space defined by the training set may not have a robust basis for prediction, for example, either returning the average of the training set or a near-random result.

Therefore, it is important to sufficiently cover chemical space such that the probability of an out-of-sample compound significantly differs from the networks internal representation of structure/property is minimized. Given estimates of up to 10^{60} potential unique structures, and that the union of all public databases amounts to only a few billion unique structures, complete chemical space coverage with current chemical knowledge cannot be guaranteed.[520−523] However, a wide range of properties, both empirical and qualitative, can be sampled to ensure coverage of descriptors used to define a chemical subspace. In deep learning applications in property prediction,[506−516] data are curated to represent a wide sampling of chemical space, relevant to the domain of application, and additional measures (cross validation, checkpointing, early stopping, learning rate decay) are taken to minimize overfitting, together maximizing model generality. As a result, the validation error, the error among predictions withheld for evaluation purposes, is only slightly higher than among predictions from data the model was trained on. This is promising as an initial assessment of the generality of such models, but as chemical space is further expanded and characterized, models should be continuously evaluated.

4.2.3. Computation Time. While per-structure computation time improves substantially when comparing deep learning to quantum chemical calculations, it is important to contextualize with required training times, which vary based on complexity of the considered network and the number of training examples. Other considerations, such as cascade or transfer learning configurations, additionally contribute to an increase in total train time. For example, a network with 5.6 million trainable parameters, trained on over 50 million chemical structures gleaned from several public databases, including PubChem,[524] UNPD,[525] HMDB,[526] and DSSTox,[527] and a curated data set with ∼500 experimental CCS values (metabolomics.pnnl.gov), requires approximately 123 h of training on a single Nvidia Tesla V100 GPU.

This compares first-principles approaches for CCS prediction,[497,499,500] which require on the order of hours per structure on a dual-socket Intel Haswell E5-2670v3 CPU. For example, employing the B3LYP exchange-correlation functional[528−531] and 6-31+G** basis set[532,533] for DFT optimization alone requires, on average, approximately 1 node-hour per conformation sampled. Propertion prediction methods involve 10 s to 100 s of conformations, rapidly increasing the per-molecule computation time. Of course, relaxing the level of DFT theory reduces computation time, although at a penalty to accuracy. This speedup extends to other physiochemical properties in which deep learning has been applied, *i.e.*, for NMR chemical shifts[210,243,534,535] and MS fragmentation patterns.[536] IR spectral prediction has yet to see a deep learning solution, but we would expect computation time to follow similar trends.

Thus, although training via machine learning requires substantial computational resources up front, we remember that, once trained, each structure can be processed on the order of milliseconds, motivating deep learning use when computational efficiency is of chief importance, *e.g.*, when building massive *in silico* libraries.

4.2.4. Accuracy. Among evaluated *in silico* prediction pipelines for CCS prediction, the best performing method resulted in an average unsigned error of 3.2%, with other methods achieving around 5% error when evaluated on the same structures. This error was achieved using the B3LYP exchange-correlation functional and 6-31+G** basis set. Higher levels of theory have been shown to reduce this error by up to 70%, but at a 2 orders of magnitude increase in computation time.[497] In contrast, deep learning has been employed to reach an average error of 2.4%, or a 25% reduction compared to first-principles approaches. Similarly, generative deep learning approaches for use in inverse QSPR applications, although sacrificing in property prediction accuracy, still improve over first-principles simulation with an average error of 3.0%.[505]

In applications involving NMR chemical shifts, deep,[534] and machine[243] learning approaches were able to achieve MAE of 0.37 and 0.28 ppm for ^{1}H shifts, respectively, and 3.3 and 3.9 ppm for ^{13}C shifts, respectively. These methods have not yet reached the accuracy provided by quantum chemical calculations, wherein MAE for ^{1}H and ^{13}C predictions can be <0.1 and <1 ppm, respectively.[204]

With respect to mass fragmentation pattern prediction, most machine learning solutions have been developed for peptide fragment prediction and produce high accuracy results (up to 0.99 Pearson correlation coefficient with experimental spectra).[537−541] Of course, fragmentation pattern prediction of metabolites represents additional challenges, and these results are thus not directly comparable. Unfortunately, the number of deep or machine learning approaches to metabolite fragmentation pattern prediction is limited. Brouard *et al.* implemented an input/output kernel regression technique and compared to a commonly used fragment prediction tool: CSI:FingerID.[536] Results indicated minor improvements over existing techniques, representing a 1−2% increase in top-*k* accuracy (37.8, 69.7, and 78.4% *versus* 36.0, 67.5, 76.5% for top 1, 5, and 10 accuracy, respectively, in positive ionization mode).

4.2.5. Inverse QSPR. An advantage to deep learning approaches, specifically those involving generative models, is their usefulness in inverse QSPR applications. The inverse QSPR problem involves determining a putative structure (or structures) based on chemical properties, for example, determining the structures that correspond to a given experimental *m*/*z* and some combination of CCS, NMR chemical shifts, IR spectra, and/or mass fragmentation pattern. With a variational autoencoder, for example, the network learns a continuous numerical, or latent, representation of the modeled structure.[542] Those dimensions of the latent representation determined to be uncorrelated with the properties of interest can be traversed to yield putative structures. Thus, the uncorrelated dimensions are varied, yet those correlated with the properties of interest remain largely invariant.

Under this framework, experimental features can be used to identify matching in-sample structures within a tolerance. These seed structures can then be perturbed as discussed to yield new structures with matching experimental signatures. This method does not solve the inverse QSPR problem because there is no guarantee the "true" structure will be generated. However, this method offers the ability to generate putative structures previously limited to those found in databases for use in high throughput virtual screening and similar approaches.

4.2.6. Interplay. Importantly, first-principles simulations and deep learning models can be leveraged for mutual benefit. Often, experimentally derived chemical properties of prediction interest are limited; for example, the union of all publicly available experimental CCS values contains only a few thousand unique structures. Training on such a limited data set, particularly when employing a large neural network, results in susceptibility to overfitting effects and, by extension, negatively affects generality. Instead, first-principles approaches can be used to expand the number of training labels, albeit with associated error, to gain broader chemical space coverage. Training can then be performed on the *in silico* values initially, followed by a round of experimental "fine tuning" to correct for the inherent error modeled from the *in silico* data. This allows for training on significantly larger data sets than would be possible with experimental values alone, the model thus learning from a broader sampling of chemical space.

Equally useful is the use of deep learning in an *in silico* property prediction pipeline, particularly used to replace computationally expensive quantum chemical calculations, *e.g.*, DFT. Under this paradigm, quantum chemical simulations are used to solve electronic structures of molecules for use in a training set. Deep learning is then leveraged to predict electronic structures from representations of atom types, positions, and chemical environments amenable to consumption by a neural network. Although data sets have thus far been fairly limited, in both size and complexity, initial results show promise with MAE less than 1%.[502]

In addition, a cascade of techniques can be combined for virtual high-throughput screening, *i.e.*, initially down-select chemical space based on machine-learning-predicted properties, verify putative structures and associated properties with quantum mechanical simulations, and confirm by experimentation. This paradigm represents a path from low cost, high throughput to high cost, low throughput, enabling researchers to efficiently use both computational and experimental resources.

AUTHOR INFORMATION

Corresponding Author

Ryan S. Renslow − *Biological Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0002-3969-5570; Email: ryan.renslow@pnnl.gov

Authors

Ricardo M. Borges − *Walter Mors Institute of Research on Natural Products, Federal University of Rio de Janeiro, Rio de Janeiro 21941-901, Brazil;* orcid.org/0000-0002-7662-6734

Sean M. Colby − *Biological Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0002-3193-8267

Susanta Das − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0001-7981-5162

Arthur S. Edison − *Departments of Genetics and Biochemistry and Molecular Biology, Complex Carbohydrate Research Center and Institute of Bioinformatics, University of Georgia, Athens, Georgia 30602, United States;* orcid.org/0000-0002-5686-2350

Oliver Fiehn − *West Coast Metabolomics Center for Compound Identification, UC Davis Genome Center, University of California, Davis, California 95616, United States;* orcid.org/0000-0002-6261-8928

Tobias Kind − *West Coast Metabolomics Center for Compound Identification, UC Davis Genome Center, University of California, Davis, California 95616, United States;* orcid.org/0000-0002-1908-4916

Jesi Lee − *West Coast Metabolomics Center for Compound Identification, UC Davis Genome Center and Department of Chemistry, University of California, Davis, California 95616, United States;* orcid.org/0000-0002-2063-4743

Amy T. Merrill − *Department of Chemistry, University of California, Davis, California 95616, United States;* orcid.org/0000-0003-4801-1721

Kenneth M. Merz, Jr. − *Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, United States;* orcid.org/0000-0001-9139-5893

Thomas O. Metz − *Biological Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0001-6049-3968

Jamie R. Nunez − *Biological Science Division, Pacific Northwest National Laboratory, Richland, Washington 99352, United States;* orcid.org/0000-0002-8594-1648

Dean J. Tantillo − *Department of Chemistry, University of California, Davis, California 95616, United States;* orcid.org/0000-0002-2992-8844

Lee-Ping Wang − *Department of Chemistry, University of California, Davis, California 95616, United States*

Shunyang Wang − *West Coast Metabolomics Center for Compound Identification, UC Davis Genome Center and Department of Chemistry, University of California, Davis, California 95616, United States;* orcid.org/0000-0003-3072-9946

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrev.0c00901

Notes

The authors declare no competing financial interest.

Biographies

Ricardo M. Borges completed his B.S. degree in pharmacy from the Federal University of Rio de Janeiro (UFRJ, Brazil) in 2004 and M.S. and Ph.D. degrees from the Walter Mors Institute of Research on Natural Products (IPPN) at the same university in 2006 and 2010. During this period, he was focusing his research on the isolation and structure elucidation methods such as in classic natural products research. Today, as PI in the same IPPN-UFRJ since 2012, his main effort is on the development of approaches for compound identification in mixtures with natural products metabolomics combining the use of MS and NMR.

Sean M. Colby is a Data Scientist in the Earth and Biological Sciences Division at Pacific Northwest National Laboratory. He received a master's degree in computer science from the Georgia Institute of Technology. His research involves *in silico* predictions of physiochemical properties by both quantum mechanical simulation and machine learning, as well as the cheminformatics pipeline metabolomic analysis relies upon. His effort has led to the development of ISiCLE, the *in silico* chemical library engine, DarkChem, a deep learning framework for molecular property prediction, and DEIMoS, data extraction in integrated multidimension spectrometry.

Susanta Das received a Ph.D. in 2015 from CSIR-National Chemical Laboratory, India (with Dr. Sourav Pal). He joined Bar-Ilan University, Israel in 2015 for postdoctoral work with Dr. Dan T. Major, where he focused on development of QM/MM simulation strategies, protein−ligand docking protocol (*viz.* EnzyDock), and enzyme catalysis applying QM/MM methods. He moved to USA in 2019 and start working with Dr. Kenneth M. Merz Jr. at Michigan State University as a postdoctoral researcher. His research with Merz group focused on *in silico* NMR and MS-based protocol development for structure identification of metabolites using FF, ML, and high-level quantum mechanical (QM) method.

Arthur S. Edison is a Georgia Research Alliance Eminent Scholar and Professor of Genetics and Biochemistry and a member of the Complex Carbohydrate Research Center and Institute of Bioinformatics at the University of Georgia. He received his Ph.D. in biophysics from the University of Wisconsin−Madison under joint supervision of John Markley and Frank Weinhold. He joined the faculty at the University of Florida and the National High Magnetic Field Laboratory in 1996. He was the founding PI and Director of the NIH-funded Southeast Center for Integrated Metabolomics. In 2015, Prof. Edison moved to the University of Georgia, where he directs the CCRC NMR facility, which supports research in both metabolomics and structural biology. Edison's research group collaborates on several metabolomics projects from microbes to humans.

Oliver Fiehn has pioneered developments and applications in metabolomics with over 320 publications to date, starting in 2000 as group leader at the Max-Planck Institute in Potsdam, Germany. Since 2004, he has been a professor at the University of California, Davis. Since 2012, he has served as Director of the NIH West Coast Metabolomics Center, overseeing his research laboratory and the satellite core service with 35 staff and 17 mass spectrometers. Professor Fiehn's research laboratory develops and implements new analytical technologies and databases for covering the metabolome. He studies fundamental biochemical questions from metabolite damage repair to epimetabolites in human diseases and animal models.

Tobias Kind is an Associate Project Scientist at the West Coast Metabolomics Center and UC Davis Genome Center. His research projects focus on the advancement of structure elucidation techniques and databases for small molecule identifications. His research interests include computational metabolomics and MS, cheminformatics, machine learning, and deep learning as well as quantum chemistry for the creation of *in silico* spectra.

Jesi Lee is a Ph.D. student under the supervision of Dr. Lee-Ping Wang and Dr. Oliver Fiehn in the Chemistry Department at the University of California, Davis. She is interested in using *ab initio* and semiempirical molecular dynamics to simulate collision-induced dissociation for improving *in silico* MS libraries.

Amy T. Merrill obtained her B.S. in chemistry at California Polytechnic State University San Luis Obispo in 2015, where she researched alkoxyamine and oxime ether synthesis with Professor Hasan Palandoken. She is currently a Ph.D. candidate with Professor Dean J. Tantillo. Her dissertation research focuses on structure elucidation of natural products and method development in computational NMR.

Kenneth M. Merz, Jr., is the Joseph Zichis Chair in Chemistry at Michigan State University and is the Editor-in-Chief of the *Journal of Chemical Information and Modeling*. His research interests lie in the development of theoretical and computational tools and their application to chemical and biological problems including structure and ligand-based drug design, mechanistic enzymology, transition ion modeling, and methodological verification and validation. He has received numerous honors, including election as an ACS Fellow, the 2010 ACS Award for Computers in Chemical and Pharmaceutical Research, election as a fellow of the American Association for the Advancement of Science, and a John Simon Guggenheim Fellowship. When not doing science, he enjoys (re)reading the novels of Jane Austen.

Thomas O. Metz joined Pacific Northwest National Laboratory in 2003 for postdoctoral work in MS with Dr. Richard D. Smith, where he focused on metabolomics. He became Staff Scientist and a Principal Investigator in the Integrative Omics Group in 2005 and is the Metabolomics Team Lead for a group of scientists that focuses on development and applications of high-throughput metabolomics and lipidomics methods to various biological questions. Dr. Metz's research has focused primarily on applying MS-based omics approaches, including proteomics, in studies of diabetes mellitus and infectious diseases, resulting in over 150 publications to date. Currently, he is the Director of the Pacific Northwest Advanced Compound Identification Core within the NIH Common Fund Metabolomics Program.

Jamie R. Nunez is working as a Data Scientist at PNNL while also pursuing a Ph.D. under the supervision of Ryan Renslow in the Chemical Engineering department at Washington State University. Her work encompasses the full metabolomics pipeline (from experimental methods to interpretation of final results) but has primarily focused on scoring for the reporting of identification results using the Multi-Attribute Matching Engine (MAME) and methods to improve confidence.

Dean J. Tantillo was born and raised in Quincy, Massachusetts, USA. He received an A.B. degree in chemistry in 1995 from Harvard and a Ph.D. in 2000 from UCLA (with Ken Houk) and then moved to Cornell, where he did postdoctoral research with Roald Hoffmann. Dean joined the faculty at UC Davis in 2003. Research in Dean's group is driven by puzzling mechanistic questions in the areas of biosynthesis, organometallic chemistry, and stereoselective synthetic reactions, with a focus on cyclization/rearrangement reactions used by Nature and by chemists to synthesize complex natural products.

Lee-Ping Wang was born and raised in the East Bay region of California and studied physics at UC Berkeley (B.A. 2006). He attended graduate school in physical chemistry at MIT advised by Troy Van Voorhis (Ph.D. 2001) and did research as a postdoc at Stanford with Todd Martinez and Vijay Pande. He joined the UC Davis Chemistry Department in 2015 and was promoted to associate professor as of July 2020. Lee-Ping's core research interests are at the intersection of quantum chemistry and molecular dynamics, where he develops methods for reaction pathway discovery and force field optimization and applies them to diverse problems in collaboration with experiment, including metabolomics.

Shunyang Wang is a doctoral student at University of California, Davis. He obtained a B.S. degree in chemistry from the Shandong University in 2018, and his doctoral work is focused on predicting mass spectra.

Ryan S. Renslow leads a team of researchers focused on computational methods to accelerate small-molecule identification in complex samples in the Biological Science Division at the Pacific Northwest National Laboratory (PNNL). He is also an associate research professor in The Gene and Linda Voiland School of Chemical Engineering and Bioengineering at Washington State University. He received his Ph.D. in chemical engineering from Washington State University in 2012, after which he joined PNNL as a Linus Pauling Distinguished Postdoctoral Fellow (2012−2015). Currently he is the colead of the Computational Core of the Pacific Northwest Advanced Compound Identification Core. His current research centers on computational quantum chemistry, machine learning and artificial intelligence, and cheminformatics methods in metabolomics, exposomics, and chemical forensics applications.

ACKNOWLEDGMENTS

This work was completed through a collaboration between the NIH-NIEHS Compound Identification Development Cores, which are part of the Common Fund Phase 2 Metabolomics Program (grants U2CES030170, U2CES030167, and U2CES030158). PNNL is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830.

ABBREVIATIONS

AIMD = *ab initio* molecular dynamics
AWS = Amazon Web Services
BOMD = Born−Oppenheimer molecular dynamics
CAD = collision-activation decomposition
CASE = computer aided structure elucidation
CCS = collision cross section
CD = circular dichroism
CE = capillary electrophoresis
CID = collision-induced dissociation
CIS = configuration interaction singles
DFT = density functional theory
EI = electron ionization
ESI = electrospray ionization
FDA = U.S. Food and Drug Administration
FF = force field
FTICR = Fourier transform ion cyclotron resonance
GC = gas chromatography
GRRM = global reaction route mapping
HMDB = Human Metabolome Database
HPC = high-performance computing
IEE = internal excess energies
IMS = ion mobility spectrometry
IP = ionization potential
IR = infrared
IRIS = infrared ion spectroscopy
IRMPD = infrared multiple-photon dissociation spectroscopy
ISiCLE = *In Silico* Chemical Library Engine
KS-DFT = Kohn−Sham density functional theory
LC = liquid chromatography
LSD = logic for structure determination
MAE = mean absolute error
MC = Monte Carlo
MD = molecular dynamics
ML = machine learning
MO = molecular orbital
MS = MS
MS/MS = tandem MS
MSI = Metabolomics Standards Initiative
NMR = nuclear magnetic resonance
PEPICO = photoelectron photoion coincidence
PES = potential energy surface
PKU = phenylketonuria
QCEIMS = quantum chemistry electron ionization MS
QET = quasi-equilibrium theory
QM = quantum mechanical
QQQ = triple-quadrupole
QSPR = quantitative structure−property relationship
Q-TOF = quadrupole-time-of-flight
RRKM = Rice−Ramsperger−Kassel−Marcus
SUMMIT = structure of unknown metabolomic mixture components by MS/NMR
TD-DFT = time-dependent density functional theory
TMS = trimethylsilyl
TOF = time of flight

REFERENCES

(1) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422*, 198−207.

(2) Giani, A. M.; Gallo, G. R.; Gianfranceschi, L.; Formenti, G. Long walk to genomics: History and current approaches to genome sequencing and assembly. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 9−19.

(3) Heather, J. M.; Chain, B. The sequence of sequencers: The history of sequencing DNA. *Genomics* **2016**, *107*, 1−8.

(4) Hashimoto, Y.; Greco, T. M.; Cristea, I. M. Contribution of mass spectrometry-based proteomics to discoveries in developmental biology. *Adv. Exp. Med. Biol.* **2019**, *1140*, 143−154.

(5) Song, Y.; Xu, X.; Wang, W.; Tian, T.; Zhu, Z.; Yang, C. Single cell transcriptomics: moving towards multi-omics. *Analyst* **2019**, *144*, 3172−3189.

(6) Nicholson, J. K.; Lindon, J. C. Systems biology: metabonomics. *Nature* **2008**, *455*, 1054−1056.

(7) Fiehn, O. Metabolomics–the link between genotypes and phenotypes. *Plant Mol. Biol.* **2002**, *48*, 155−171.

(8) Sauer, U.; Heinemann, M.; Zamboni, N. Genetics. Getting closer to the whole picture. *Science* **2007**, *316*, 550−551.

(9) Nurse, P.; Hayles, J. The cell in an era of systems biology. *Cell* **2011**, *144*, 850−854.

(10) Westerhoff, H. V.; Palsson, B. O. The evolution of molecular biology into systems biology. *Nat. Biotechnol.* **2004**, *22*, 1249−1252.

(11) Cuccato, G.; Della Gatta, G.; di Bernardo, D. Systems and synthetic biology: tackling genetic networks and complex diseases. *Heredity* **2009**, *102*, 527−532.

(12) Fletcher, D. Which biological systems should be engineered? *Nature* **2018**, *563*, 177−179.

(13) Yeh, B. J.; Lim, W. A. Synthetic biology: lessons from the history of synthetic organic chemistry. *Nat. Chem. Biol.* **2007**, *3*, 521−525.

(14) Djurdjevic, L.; Mitrovic, M.; Pavlovic, P. Methodology of allelopathy research: 2. forest ecosystems. *Allelopathy J.* **2007**, *20*, 79−102.

(15) Li, J. W. H.; Vederas, J. C. Drug discovery and natural products: end of an era or an endless frontier? *Science* **2009**, *325*, 161−165.

(16) Williams, D. H.; Stone, M. J.; Hauck, P. R.; Rahman, S. K. Why are secondary metabolites (natural-products) biosynthesized. *J. Nat. Prod.* **1989**, *52*, 1189−1208.

(17) Bennett, R. N.; Wallsgrove, R. M. Secondary metabolites in plant defense-mechanisms. *New Phytol.* **1994**, *127*, 617−633.

(18) Li, Z. H.; Wang, Q. A.; Ruan, X. A.; Pan, C. D.; Jiang, D. A. Phenolics and Plant Allelopathy. *Molecules* **2010**, *15*, 8933−8952.

(19) Vyvyan, J. R. Allelochemicals as leads for new herbicides and agrochemicals. *Tetrahedron* **2002**, *58*, 1631−1646.

(20) Wardle, D. A.; Nilsson, M. C.; Gallet, C.; Zackrisson, O. An ecosystem-level perspective of allelopathy. *Biol. Rev.* **1998**, *73*, 305−319.

(21) Easton, D. F.; Bishop, D. T.; Ford, D.; Crockford, G. P. Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The breast cancer linkage consortium. *Am. J. Hum. Genet.* **1993**, *52*, 678−701.

(22) Hall, J. M.; Lee, M. K.; Newman, B.; Morrow, J. E.; Anderson, L. A.; Huey, B.; King, M. C. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **1990**, *250*, 1684−1689.

(23) Narod, S. A.; Lynch, H. T.; Watson, P.; Conway, T.; Lynch, J.; Lenoir, G. M.; Feunteun, S. Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet* **1991**, *338*, 82−83.

(24) Miki, Y.; Swensen, J.; Shattuck-Eidens, D.; Futreal, P. A.; Harshman, K.; Tavtigian, S.; Liu, Q.; Cochran, C.; Bennett, L. M.; Ding, W.; et al. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **1994**, *266*, 66−71.

(25) Wooster, R.; Neuhausen, S. L.; Mangion, J.; Quirk, Y.; Ford, D.; Collins, N.; Nguyen, K.; Seal, S.; Tran, T.; Averill, D.; et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12−13. *Science* **1994**, *265*, 2088−2090.

(26) Kuchenbaecker, K. B.; Hopper, J. L.; Barnes, D. R.; Phillips, K. A.; Mooij, T. M.; Roos-Blom, M. J.; Jervis, S.; van Leeuwen, F. E.; Milne, R. L.; Andrieu, N.; et al. Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. *JAMA* **2017**, *317*, 2402−2416.

(27) Zhang, Z.; Bast, R. C., Jr; Yu, Y.; Li, J.; Sokoll, L. J.; Rai, A. J.; Rosenzweig, J. M.; Cameron, B.; Wang, Y. Y.; Meng, X. Y.; et al. Three biomarkers identified from serum proteomic analysis for the detection of early stage ovarian cancer. *Cancer Res.* **2004**, *64*, 5882−5890.

(28) Metz, T. O.; Baker, E. S.; Schymanski, E. L.; Renslow, R. S.; Thomas, D. G.; Causon, T. J.; Webb, I. K.; Hann, S.; Smith, R. D.; Teeguarden, J. G. Integrating ion mobility spectrometry into mass spectrometry-based exposome measurements: what can it add and how far can it go? *Bioanalysis* **2017**, *9*, 81−98.

(29) Jansson, J. K.; Baker, E. S. A multi-omic future for microbiome studies. *Nature Microbiology* **2016**, *1*, 16049.

(30) Rashed, M. S. Clinical applications of tandem mass spectrometry: ten years of diagnosis and screening for inherited metabolic diseases. *J. Chromatogr., Biomed. Appl.* **2001**, *758*, 27−48.

(31) Clayton, P. T. Applications of mass spectrometry in the study of inborn errors of metabolism. *J. Inherited Metab. Dis.* **2001**, *24*, 139−150.

(32) Guthrie, R.; Susi, A. A Simple phenylalanine method for detecting phenylketonuria in large populations of newborn infants. *Pediatrics* **1963**, *32*, 338−343.

(33) Berry, S. A.; Brown, C.; Grant, M.; Greene, C. L.; Jurecki, E.; Koch, J.; Moseley, K.; Suter, R.; van Calcar, S. C.; Wiles, J.; Cederbaum, S. Newborn screening 50 years later: access issues faced by adults with PKU. *Genet. Med.* **2013**, *15*, 591−599.

(34) Watson, J. D.; Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **1953**, *171*, 737−738.

(35) Franklin, R. E.; Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **1953**, *171*, 740−741.

(36) Wilkins, M. H.; Stokes, A. R.; Wilson, H. R. Molecular structure of deoxypentose nucleic acids. *Nature* **1953**, *171*, 738−740.

(37) Lehman, I. R.; Bessman, M. J.; Simms, E. S.; Kornberg, A. Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli. *J. Biol. Chem.* **1958**, *233*, 163−170.

(38) Kleppe, K.; Ohtsuka, E.; Kleppe, R.; Molineux, I.; Khorana, H. G. Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. *J. Mol. Biol.* **1971**, *56*, 341−361.

(39) Mullis, K. B. The unusual origin of the polymerase chain reaction. *Sci. Am.* **1990**, *262* (56−61), 64−55.

(40) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.

(41) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551−3567.

(42) Ma, K.; Vitek, O.; Nesvizhskii, A. I. A statistical model-building perspective to identification of MS/MS spectra with PeptideProphet. *BMC Bioinf.* **2012**, *13* (16), S1.

(43) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* **2010**, *604*, 55−71.

(44) Metz, T. O.; Zhang, Q.; Page, J. S.; Shen, Y.; Callister, S. J.; Jacobs, J. M.; Smith, R. D. The future of liquid chromatography-mass spectrometry (LC-MS) in metabolic profiling and metabolomic studies for biomarker discovery. *Biomarkers Med.* **2007**, *1*, 159−185.

(45) Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* **1996**, *16*, 3−50.

(46) Polishchuk, P. G.; Madzhidov, T. I.; Varnek, A. Estimation of the size of drug-like chemical space based on GDB-17 data. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 675−679.

(47) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinf.* **2006**, *7*, 234.

(48) Wishart, D. S.; Knox, C.; Guo, A. C.; Eisner, R.; Young, N.; Gautam, B.; Hau, D. D.; Psychogios, N.; Dong, E.; Bouatra, S.; et al. HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* **2009**, *37*, D603−610.

(49) DSSTox, E...

(50) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J.; ACS Publications, 2014.

(51) Palermo, A. Charting metabolism heterogeneity by nanostructure imaging mass spectrometry: from biological systems to subcellular functions. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 2392.

(52) Heyman, H. M.; Dubery, I. A. The potential of mass spectrometry imaging in plant metabolomics: a review. *Phytochem. Rev.* **2016**, *15*, 297−316.

(53) Miura, D.; Fujimura, Y.; Wariishi, H. In situ metabolomic mass spectrometry imaging: Recent advances and difficulties. *J. Proteomics* **2012**, *75*, 5052−5060.

(54) Palmer, A.; Phapale, P.; Chernyavsky, I.; Lavigne, R.; Fay, D.; Tarasov, A.; Kovalev, V.; Fuchser, J.; Nikolenko, S.; Pineau, C.; Becker, M.; Alexandrov, T. FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry. *Nat. Methods* **2017**, *14*, 57−60.

(55) Butler, M. C.; Mehta, H. S.; Chen, Y.; Reardon, P. N.; Renslow, R. S.; Khbeis, M.; Irish, D.; Mueller, K. T. Toward high-resolution NMR spectroscopy of microscopic liquid samples. *Phys. Chem. Chem. Phys.* **2017**, *19*, 14256−14261.

(56) Chen, Y.; Mehta, H. S.; Butler, M. C.; Walter, E. D.; Reardon, P. N.; Renslow, R. S.; Mueller, K. T.; Washton, N. M. High-resolution microstrip NMR detectors for subnanoliter samples. *Phys. Chem. Chem. Phys.* **2017**, *19*, 28163−28174.

(57) Ramaswamy, V.; Hooker, J. W.; Withers, R. S.; Nast, R. E.; Brey, W. W.; Edison, A. S. Development of a C-13-optimized 1.5-mm high temperature superconducting NMR probe. *J. Magn. Reson.* **2013**, *235*, 58−65.

(58) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic H-1 NMR data sets. *Anal. Chem.* **2005**, *77*, 1282−1289.

(59) Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. Evaluation of the orthogonal projection on latent structure model limitations caused by chemical shift variability and improved visualization of biomarker changes in H-1 NMR spectroscopic metabonomic studies. *Anal. Chem.* **2005**, *77*, 517−526.

(60) Robinette, S. L.; Brüschweiler, R.; Schroeder, F. C.; Edison, A. S. NMR in metabolomics and natural products research: two sides of the same coin. *Acc. Chem. Res.* **2012**, *45*, 288−297.

(61) Jones, C. G.; Martynowycz, M. W.; Hattne, J.; Fulton, T. J.; Stoltz, B. M.; Rodriguez, J. A.; Nelson, H. M.; Gonen, T. The CryoEM method microED as a powerful tool for small molecule structure determination. *ACS Cent. Sci.* **2018**, *4*, 1587−1592.

(62) Castle, A. L.; Fiehn, O.; Kaddurah-Daouk, R.; Lindon, J. C. Metabolomics standards workshop and the development of international standards for reporting metabolomics experimental results. *Briefings Bioinf.* **2006**, *7*, 159−165.

(63) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* **2007**, *3*, 211−221.

(64) Beisken, S.; Eiden, M.; Salek, R. M. Getting the right answers: understanding metabolomics challenges. *Expert Rev. Mol. Diagn.* **2015**, *15*, 97−109.

(65) Markley, J. L.; Brüschweiler, R.; Edison, A. S.; Eghbalnia, H. R.; Powers, R.; Raftery, D.; Wishart, D. S. The future of NMR-based metabolomics. *Curr. Opin. Biotechnol.* **2017**, *43*, 34−40.

(66) Tulp, M.; Bohlin, L. Functional versus chemical diversity: is biodiversity important for drug discovery? *Trends Pharmacol. Sci.* **2002**, *23*, 225−231.

(67) Schymanski, E. L.; Singer, H. P.; Slobodnik, J.; Ipolyi, I. M.; Oswald, P.; Krauss, M.; Schulze, T.; Haglund, P.; Letzel, T.; Grosse,

S.; et al. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem.* **2015**, *407*, 6237−6255.

(68) Levine, B. G.; Martinez, T. J. Isomerization through conical intersections. *Annu. Rev. Phys. Chem.* **2007**, *58*, 613−634.

(69) Marian, C. M. Spin-orbit coupling and intersystem crossing in molecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *2*, 187−203.

(70) Hartree, D. R. The wave mechanics of an atom with a non-Coulomb central field Part I theory and methods. *Math. Proc. Cambridge Philos. Soc.* **1928**, *24*, 89−110.

(71) Fock, V. Approximation method for the solution of the quantum mechanical multibody problems. *Eur. Phys. J. A* **1930**, *61*, 126−148.

(72) Thomas, L. H. The calculation of atomic fields. *Math. Proc. Cambridge Philos. Soc.* **1927**, *23*, 542−548.

(73) Fermi, E. A statistical Method for Determining some Properties of the Atoms and its Application to the Theory of the periodic Table of Elements. *Eur. Phys. J. A* **1928**, *48*, 73−79.

(74) Dirac, P. A. M. The quantum theory of the electron. *Proc. R. Soc. London Ser. A- Contain. Pap. Math. Phys. Character* **1928**, *117*, 610−624.

(75) Pople, J.; Santry, D.; Segal, G. Approximate Self-Consistent Molecular Orbital Theory.i. Invariant Procedures. *J. Chem. Phys.* **1965**, *43*, S129−S135.

(76) Hehre, W.; Stewart, R.; Pople, J. Self-Consistent Molecular-Orbital Methods.i. Use of Gaussian Expansions. *J. Chem. Phys.* **1969**, *51*, 2657−2664.

(77) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864−B871.

(78) Kohn, W.; Sham, L. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.

(79) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16*, Revision C.01; Gaussian, Inc.: Wallingford CT, 2016..

(80) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **2015**, *113*, 184−215.

(81) Gordon, M. S.; Schmidt, M. W. *Advances in electronic structure theory: GAMESS a decade later*: Amsterdam, 2005.

(82) Parrish, R. M.; Burns, L. A.; Smith, D. G. A.; Simmonett, A. C.; DePrince, A. E.; Hohenstein, E. G.; Bozkaya, U.; Sokolov, A. Y.; Di Remigio, R.; Richard, R. M.; et al. PSI4 1.1: An Open-Source Electronic Structure Program Emphasizing Automation, Advanced Libraries, and Interoperability. *J. Chem. Theory Comput.* **2017**, *13*, 3185−3197.

(83) Werner, H. J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M.; Celani, P.; Györffy, W.; Kats, D.; Korona, T.; Lindh, R.; et al. *MOLPRO, Version 2019.2, a Package of ab Initio Programs*: Cardiff University: Cardiff, UK, 2019.

(84) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L.; de Jong, W. NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477−1489.

(85) Neese, F. Software update: the ORCA program system, version 4.0. *WIREs Computational Molecular Science* **2018**, *8*, No. e1327.

(86) Ufimtsev, I. S.; Martinez, T. J. Quantum chemistry on graphical processing units. 3. analytical Energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619−2628.

(87) Titov, A. V.; Ufimtsev, I. S.; Luehr, N.; Martinez, T. J. Generating efficient quantum chemistry codes for novel architectures. *J. Chem. Theory Comput.* **2013**, *9*, 213−221.

(88) Pople, J. A. Nobel Lecture: Quantum chemical models. *Rev. Mod. Phys.* **1999**, *71*, 1267−1274.

(89) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. Gaussian-2 theory for molecular energies of first- and second-row compounds. *J. Chem. Phys.* **1991**, *94*, 7221−7230.

(90) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. Assessment of Gaussian-3 and density-functional theories on the G3/05 test set of experimental energies. *J. Chem. Phys.* **2005**, *123*, 124107.

(91) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479−483.

(92) Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga, J. Non-iterative fifth-order triple and quadruple excitation energy corrections in correlated methods. *Chem. Phys. Lett.* **1990**, *165*, 513−522.

(93) Miliordos, E.; Xantheas, S. S. An accurate and efficient computational protocol for obtaining the complete basis set limits of the binding energies of water clusters at the MP2 and CCSD(T) levels of theory: Application to (H2O)m, m = 2−6, 8, 11, 16, and 17. *J. Chem. Phys.* **2015**, *142*, 234303.

(94) Feller, D.; Peterson, K. A.; Hill, J. G. On the effectiveness of CCSD(T) complete basis set extrapolations for atomization energies. *J. Chem. Phys.* **2011**, *135*, 044102.

(95) Karton, A.; Martin, J. M. L. Explicitly correlated W *n* theory: W1-F12 and W2-F12. *J. Chem. Phys.* **2012**, *136*, 124114.

(96) Roos, B.; Taylor, P.; Siegbahn, P. A Complete active space scf method (casscf) using a density-matrix formulated super-Ci approach. *Chem. Phys.* **1980**, *48*, 157−173.

(97) Andersson, K.; Malmqvist, P.; Roos, B. 2nd-order perturbation-theory with a complete active space self-consistent field reference function. *J. Chem. Phys.* **1992**, *96*, 1218−1226.

(98) Hirao, K. Multireference moller-plesset method. *Chem. Phys. Lett.* **1992**, *190*, 374−380.

(99) Höfener, S.; Klopper, W. Analytical nuclear gradients of the explicitly correlated Møller-Plesset second-order energy. *Mol. Phys.* **2010**, *108*, 1783−1796.

(100) Adamowicz, L.; Laidig, W. D.; Bartlett, R. J. Analytical gradients for the coupled-cluster method. *Int. J. Quantum Chem.* **1984**, *26*, 245−254.

(101) Hickey, A. L.; Rowley, C. N. Benchmarking quantum chemical methods for the calculation of molecular dipole moments and polarizabilities. *J. Phys. Chem. A* **2014**, *118*, 3678−3687.

(102) Baik, M. H.; Friesner, R. A. Computing redox potentials in solution: density functional theory as a tool for rational design of redox agents. *J. Phys. Chem. A* **2002**, *106*, 7407−7412.

(103) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *J. Phys. Chem. B* **2009**, *113*, 1253−1272.

(104) Li, G.; Zhang, X.; Cui, Q. Free energy perturbation calculations with combined QM/MM potentials complications, simplifications, and applications to redox potential calculations. *J. Phys. Chem. B* **2003**, *107*, 8643−8653.

(105) Brooks, B.; Janezic, D.; Karplus, M. Harmonic-Analysis of large systems.1. methodology. *J. Comput. Chem.* **1995**, *16*, 1522−1542.

(106) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Use of Solution-phase vibrational frequencies in continuum models for the free energy of solvation. *J. Phys. Chem. B* **2011**, *115*, 14556−14562.

(107) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646−652.

(108) Hollingsworth, S. A.; Dror, R. O. Molecular dynamics simulation for all. *Neuron* **2018**, *99*, 1129−1143.

(109) Paquet, E.; Viktor, H. L. Molecular dynamics, monte carlo simulations, and Langevin dynamics: a computational review. *BioMed Res. Int.* **2015**, *2015*, 183918.

(110) Bunfield, D. H.; Davis, L. M. Monte Carlo simulation of a single-molecule detection experiment. *Appl. Opt.* **1998**, *37*, 2315−2326.

(111) Ojeda, P.; Garcia, M. E.; Londoño, A.; Chen, N.-Y. Monte Carlo simulations of proteins in cages: Influence of confinement on the stability of intermediate states. *Biophys. J.* **2009**, *96*, 1076−1082.
(112) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3093.
(113) Tuckerman, M.; Berne, B.; Martyna, G.; Klein, M. Efficient molecular-dynamics and hybrid Monte-Carlo algorithms for path-integrals. *J. Chem. Phys.* **1993**, *99*, 2796−2808.
(114) Cao, J.; Voth, G. The formulation of quantum-statistical mechanics based on the feynman path centroid density.2. dynamical properties. *J. Chem. Phys.* **1994**, *100*, 5106−5117.
(115) Craig, I. R.; Manolopoulos, D. E. Quantum statistics and classical mechanics: Real time correlation functions from ring polymer molecular dynamics. *J. Chem. Phys.* **2004**, *121*, 3368−3373.
(116) Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions - dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227−249.
(117) Field, M.; Bash, P.; Karplus, M. A Combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.
(118) Al-Mogren, M. M.; El-Gogary, T. M. Structure, stability, energy barrier and ionization energies of chemically modified DNA-bases: Quantum chemical calculations on 37 favored and rare tautomeric forms of tetraphosphoadenine. *Comput. Theor. Chem.* **2015**, *1052*, 35−41.
(119) Crespo-Hernández, C. E.; Arce, R.; Ishikawa, Y.; Gorb, L.; Leszczynski, J.; Close, D. M. Ab initio ionization energy thresholds of DNA and RNA bases in gas phase and in aqueous solution. *J. Phys. Chem. A* **2004**, *108*, 6373−6377.
(120) Range, K.; Riccardi, D.; Cui, Q.; Elstner, M.; York, D. M. Benchmark calculations of proton affinities and gas-phase basicities of molecules important in the study of biological phosphoryl transfer. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3070−3079.
(121) Raabe, G.; Wang, Y. K.; Fleischhauer, J. Calculation of the proton affinities of primary, secondary, and tertiary amines using semiempirical and ab initio methods. *Z. Naturforsch., A: Phys. Sci.* **2000**, *55*, 687−694.
(122) Jursic, B. S.; Martin, R. M. Calculation of bond dissociation energies for oxygen containing molecules by ab initio and density functional theory methods. *Int. J. Quantum Chem.* **1996**, *59*, 495−501.
(123) Schroder, S.; Buckley, N.; Oppenheimer, N.; Kollman, P. A Quantum chemical study of the Type-Iv nucleophilic-substitution reaction and dissociation of the beta-nicotinamide glycosyl bond in the gas-phase using semiempirical Pm3 calculations. *J. Am. Chem. Soc.* **1992**, *114*, 8232−8238.
(124) Sengupta, D.; Chandra, A. Role of the Hno3[-]noh isomerization in reactions (i) Nh((3)sigma(−))+o(p-3) and (ii) N(s-4)+oh((2)pi) - Ab-initio calculations and quantum-statistical rice-ramsperger-kassel analysis of the potential-energy surfaces. *J. Chem. Phys.* **1994**, *101*, 3906−3915.
(125) Ha, T. K.; Keller, H. J.; Gunde, R.; Gunthard, H. H. Energy increment method based on quantum chemical results: A general recipe for approximative prediction of isomerization and tautomerization energies of pyrimidine and purine nucleic acid bases and related compounds. *J. Phys. Chem. A* **1999**, *103*, 6612−6623.
(126) Grimme, S.; Steinmetz, M.; Korth, M. How to compute isomerization energies of organic molecules with quantum chemical methods. *J. Org. Chem.* **2007**, *72*, 2118−2126.
(127) Altun, A.; Neese, F.; Bistoni, G. HFLD: A nonempirical London dispersion-corrected Hartree-Fock method for the quantification and analysis of noncovalent interaction energies of large molecular systems. *J. Chem. Theory Comput.* **2019**, *15*, 5894−5907.
(128) Toupkanloo, H. A.; Rahmani, Z. An in-depth study on noncovalent stacking interactions between DNA bases and aromatic drug fragments using DFT method and AIM analysis: conformers, binding energies, and charge transfer. *Appl. Biol. Chem.* **2018**, *61*, 209−226.
(129) Zhao, Y.; Truhlar, D. G. Density functionals for noncovalent interaction energies of biological importance. *J. Chem. Theory Comput.* **2007**, *3*, 289−300.
(130) Beachy, M. D.; Chasman, D.; Murphy, R. B.; Halgren, T. A.; Friesner, R. A. Accurate ab initio quantum chemical determination of the relative energetics of peptide conformations and assessment of empirical force fields. *J. Am. Chem. Soc.* **1997**, *119*, 5908−5920.
(131) Csaszar, A. G.; Allen, W. D.; Schaefer, H. F. In pursuit of the ab initio limit for conformational energy prototypes. *J. Chem. Phys.* **1998**, *108*, 9751−9764.
(132) Headgordon, T.; Headgordon, M.; Frisch, M.; Brooks, C.; Pople, J. Theoretical-study of blocked glycine and alanine peptide Analogs. *J. Am. Chem. Soc.* **1991**, *113*, 5989−5997.
(133) Jacquemin, D.; Wathelet, V.; Perpète, E. A.; Adamo, C. Extensive TD-DFT benchmark: singlet-excited states of Organic molecules. *J. Chem. Theory Comput.* **2009**, *5*, 2420−2435.
(134) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670−6688.
(135) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.
(136) Rezac, J. Cuby: An integrative framework for computational chemistry. *J. Comput. Chem.* **2016**, *37*, 1230−1237.
(137) Frewen, B. E.; Merrihew, G. E.; Wu, C. C.; Noble, W. S.; MacCoss, M. J. Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal. Chem.* **2006**, *78*, 5678−5684.
(138) Benton, H. P.; Wong, D. M.; Trauger, S. A.; Siuzdak, G. XCMS2: Processing tandem mass spectrometry data for metabolite identification and structural characterization. *Anal. Chem.* **2008**, *80*, 6382−6389.
(139) Wolf, S.; Schmidt, S.; Mueller-Hannemann, M.; Neumann, S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinf.* **2010**, *11*, 148.
(140) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Benton, H. P.; Siuzdak, G. Annotation: A Computational solution for streamlining metabolomics analysis. *Anal. Chem.* **2018**, *90*, 480−489.
(141) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(142) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(143) Tkatchenko, A.; Scheffler, M. Accurate molecular Van Der Waals interactions from ground-state electron density and free-atom reference data. *Phys. Rev. Lett.* **2009**, *102*, 073005.
(144) Grossert, J. S.; Cubero Herrera, L.; Ramaley, L.; Melanson, J. E. Studying the chemistry of cationized triacylglycerols using electrospray ionization mass spectrometry and density functional theory computations. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 1421−1440.
(145) Kruve, A.; Kaupmees, K. Adduct formation in ESI/MS by mobile phase additives. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 887−894.
(146) Zheng, Y.; Jiao, Y.; Jaroniec, M.; Qiao, S. Z. Advancing the Electrochemistry of the hydrogen-evolution reaction through combining experiment and theory. *Angew. Chem., Int. Ed.* **2015**, *54*, 52−65.
(147) Zhang, L.; Xia, Z. Mechanisms of oxygen reduction reaction on nitrogen-doped graphene for fuel cells. *J. Phys. Chem. C* **2011**, *115*, 11170−11176.
(148) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. Theoretical perspective on the structure and mechanism of cytochrome P450 enzymes. *Chem. Rev.* **2005**, *105*, 2279−2328.
(149) Bligaard, T.; Norskov, J. K.; Dahl, S.; Matthiesen, J.; Christensen, C. H.; Sehested, J. The Bronsted-Evans-Polanyi relation

and the volcano curve in heterogeneous catalysis. *J. Catal.* **2004**, *224*, 206−217.

(150) Balcells, D.; Clot, E.; Eisenstein, O. C-H bond activation in transition metal species from a computational perspective. *Chem. Rev.* **2010**, *110*, 749−823.

(151) Ōki, M. In *Topics in Stereochemistry*; Allinger, N. L., Eliel, E. L., Wilen, S. H., Eds.; John Wiley & Sons, 1983; Vol. *14*, pp 1−81

(152) Jursic, B. S. *Theor. Comput. Chem.* **1996**, *4*, 709−741.

(153) Rodríguez, A. M.; Prieto, P.; de la Hoz, A.; Díaz-Ortiz, Á.; Martín, D. R.; García, J. I. Influence of polarity and activation energy in microwave-assisted organic synthesis (MAOS). *Chemistry Open* **2015**, *4*, 308−317.

(154) O'Leary, W. C.; Goddard, W. A.; Cheng, M.-J. Dual-phase mechanism for the catalytic conversion of n-Butane to maleic anhydride by the vanadyl pyrophosphate heterogeneous catalyst. *J. Phys. Chem. C* **2017**, *121*, 24069−24076.

(155) Zimmerman, P. M. Navigating molecular space for reaction mechanisms: an efficient, automated procedure. *Mol. Simul.* **2015**, *41*, 43−54.

(156) Mardirossian, N.; Head-Gordon, M. ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.* **2014**, *16*, 9904−9924.

(157) Zhao, Y.; Truhlar, D. G. Exploring the limit of accuracy of the global hybrid meta density functional for main-group Thermochemistry, kinetics, and noncovalent interactions. *J. Chem. Theory Comput.* **2008**, *4*, 1849−1868.

(158) Rezac, J.; Hobza, P. Ab Initio Quantum Mechanical Description of Noncovalent Interactions at Its Limits: Approaching the Experimental Dissociation Energy of the HF Dimer. *J. Chem. Theory Comput.* **2014**, *10*, 3066−3073.

(159) Maeda, S.; Ohno, K.; Morokuma, K. Systematic exploration of the mechanism of chemical reactions: the global reaction route mapping (GRRM) strategy using the ADDF and AFIR methods. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3683−3701.

(160) Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **2014**, *6*, 1044−1048.

(161) Bergeler, M.; Simm, G. N.; Proppe, J.; Reiher, M. Heuristics-Guided Exploration of Reaction Mechanisms. *J. Chem. Theory Comput.* **2015**, *11*, 5712−5722.

(162) Maeda, S.; Harabuchi, Y.; Takagi, M.; Taketsugu, T.; Morokuma, K. Artificial Force Induced Reaction (AFIR) Method for Exploring Quantum Chemical Potential Energy Surfaces. *Chem. Rec.* **2016**, *16*, 2232−2248.

(163) Jacobson, L. D.; Bochevarov, A. D.; Watson, M. A.; Hughes, T. F.; Rinaldo, D.; Ehrlich, S.; Steinbrecher, T. B.; Vaitheeswaran, S.; Philipp, D. M.; Halls, M. D.; Friesner, R. A. Automated Transition State Search and Its Application to Diverse Types of Organic Reactions. *J. Chem. Theory Comput.* **2017**, *13*, 5780−5797.

(164) Plehiers, P. P.; Marin, G. B.; Stevens, C. V.; Van Geem, K. M. Automated reaction database and reaction network analysis: extraction of reaction templates using cheminformatics. *J. Cheminf.* **2018**, *10*, 11.

(165) Rodriguez, A.; Rodriguez-Fernandez, R.; Vazquez, S. A.; Barnes, G. L.; Stewart, J. P.; Martinez-Nunez, E. tsscds2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics. *J. Comput. Chem.* **2018**, *39*, 1922−1930.

(166) Martinez-Nunez, E. An automated transition state search using classical trajectories initialized at multiple minima. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14912−14921.

(167) Besora, M.; Vidossich, P.; Lledos, A.; Ujaque, G.; Maseras, F. Calculation of Reaction Free Energies in Solution: A Comparison of Current Approaches. *J. Phys. Chem. A* **2018**, *122*, 1392−1399.

(168) Winget, P.; Weber, E. J.; Cramer, C. J.; Truhlar, D. G. Computational electrochemistry: aqueous one-electron oxidation potentials for substituted anilines. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1231−1239.

(169) Tomasi, J.; Persico, M. Molecular-interactions in solution - an overview of methods based on continuous distributions of the solvent. *Chem. Rev.* **1994**, *94*, 2027−2094.

(170) Marten, B.; Kim, K.; Cortis, C.; Friesner, R. A.; Murphy, R. B.; Ringnalda, M. N.; Sitkoff, D.; Honig, B. New model for calculation of solvation free energies: Correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bonding effects. *J. Phys. Chem.* **1996**, *100*, 11775−11788.

(171) Cramer, C. J.; Truhlar, D. G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161−2200.

(172) Marenich, A. V.; Ho, J.; Coote, M. L.; Cramer, C. J.; Truhlar, D. G. Computational electrochemistry: prediction of liquid-phase reduction potentials. *Phys. Chem. Chem. Phys.* **2014**, *16*, 15068−15106.

(173) Ho, J.; Coote, M. L. A universal approach for continuum solvent pK(a) calculations: are we there yet? *Theor. Chem. Acc.* **2010**, *125*, 3−21.

(174) Alongi, K. S.; Shields, G. C. In *Annual Reports in Computational Chemistry*; Wheeler, R. A., Wheeler, R. A., Eds.: Elsevier: Amsterdam, 2010; Vol. *6*, pp 113−138.

(175) Tummanapelli, A. K.; Vasudevan, S. Ab initio molecular dynamics simulations of amino acids in aqueous solutions: estimating pKa values from metadynamics sampling. *J. Phys. Chem. B* **2015**, *119*, 12249−12255.

(176) Ruiz Pestana, L.; Mardirossian, N.; Head-Gordon, M.; Head-Gordon, T. Ab initio molecular dynamics simulations of liquid water using high quality meta-GGA functionals. *Chem. Sci.* **2017**, *8*, 3554−3565.

(177) Chen, M.; Ko, H.-Y.; Remsing, R. C.; Calegari Andrade, M. F.; Santra, B.; Sun, Z.; Selloni, A.; Car, R.; Klein, M. L.; Perdew, J. P.; Wu, X. Ab initio theory and modeling of water. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 10846−10851.

(178) Wilson, H. F.; Wong, M. L.; Militzer, B. Superionic to superionic phase change in water: consequences for the interiors of Uranus and Neptune. *Phys. Rev. Lett.* **2013**, *110*, 151102.

(179) Bligh, E. G.; Dyer, W. J. A Rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **1959**, *37*, 911−917.

(180) Matyash, V.; Liebisch, G.; Kurzchalia, T. V.; Shevchenko, A.; Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *J. Lipid Res.* **2008**, *49*, 1137−1146.

(181) Nakayasu, E. S.; Nicora, C. D.; Sims, A. C.; Burnum-Johnson, K. E.; Kim, Y. M.; Kyle, J. E.; Matzke, M. M.; Shukla, A. K.; Chu, R. K.; Schepmoes, A. A.; et al. MPLEx: a Robust and universal protocol for single-sample integrative proteomic, metabolomic, and lipidomic analyses. *Msystems* **2016**, *1*, e00043-16.

(182) Bannan, C. C.; Calabró, G.; Kyu, D. Y.; Mobley, D. L. Calculating partition coefficients of small molecules in octanol/water and cyclohexane/water. *J. Chem. Theory Comput.* **2016**, *12*, 4015−4024.

(183) Taskinen, J.; Yliruusi, J. Prediction of physicochemical properties based on neural network modelling. *Adv. Drug Delivery Rev.* **2003**, *55*, 1163−1183.

(184) Vitha, M.; Carr, P. W. The chemical interpretation and practice of linear solvation energy relationships in chromatography. *Journal of Chromatography A* **2006**, *1126*, 143−194.

(185) Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. Calculation of molecular lipophilicity: state-of-the-art and comparison of log P methods on more than 96,000 compounds. *J. Pharm. Sci.* **2009**, *98*, 861−893.

(186) Merrick, J. P.; Moran, D.; Radom, L. An evaluation of harmonic vibrational frequency scale factors. *J. Phys. Chem. A* **2007**, *111*, 11683−11700.

(187) Fortenberry, R. C.; Huang, X.; Yachmenev, A.; Thiel, W.; Lee, T. J. On the use of quartic force fields in variational calculations. *Chem. Phys. Lett.* **2013**, *574*, 1−12.

(188) Adel, A.; Dennison, D. M. The infrared spectrum of carbon dioxide. Part I. *Phys. Rev.* **1933**, *43*, 716.

(189) Whitehead, R.; Handy, N. Variational calculation of vibration-rotation energy-levels for triatomic-molecules. *J. Mol. Spectrosc.* **1975**, *55*, 356−373.

(190) Kubo, R. Statistical-mechanical theory of irreversible processes. I. general theory and simple applications to magnetic and conduction problems. *J. Phys. Soc. Jpn.* **1957**, *12*, 570−586.

(191) Gordon, R. G. Molecular motion in infrared and raman spectra. *J. Chem. Phys.* **1965**, *43*, 1307−1312.

(192) Foresman, J. B.; Head-Gordon, M.; Pople, J. A.; Frisch, M. J. Toward a systematic molecular orbital theory for excited states. *J. Phys. Chem.* **1992**, *96*, 135−149.

(193) Runge, E.; Gross, E. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.* **1984**, *52*, 997−1000.

(194) Werner, H. J.; Meyer, W. A quadratically convergent MCSCF method for the simultaneous optimization of several states. *J. Chem. Phys.* **1981**, *74*, 5794−5801.

(195) Roos, B. O.; Andersson, K.; Fulscher, M. P.; Malmqvist, P. A.; Serrano Andres, L.; Pierloot, K.; Merchan, M. Multiconfigurational perturbation theory: applications in electronic spectroscopy. *Adv. Chem. Phys.* **1996**, *93*, 219−331.

(196) Elyashberg, M. Identification and structure elucidation by NMR spectroscopy. *TrAC, Trends Anal. Chem.* **2015**, *69*, 88−97.

(197) Le Guennec, A.; Tea, I.; Antheaume, I.; Martineau, E.; Charrier, B.; Pathan, M.; Akoka, S.; Giraudeau, P. Fast determination of absolute metabolite concentrations by spatially encoded 2D NMR: Application to breast cancer cell extracts. *Anal. Chem.* **2012**, *84*, 10831−10837.

(198) Coen, M.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology. *Chem. Res. Toxicol.* **2008**, *21*, 9−27.

(199) Pimenta, L. P. S.; Kim, H. K.; Verpoorte, R.; Choi, Y. H. NMR-based metabolomics: A probe to utilize biodiversity. *Methods Mol. Biol.* **2013**, *1055*, 117−127.

(200) Clendinen, C. S.; Stupp, G. S.; Wang, B.; Garrett, T. J.; Edison, A. S. 13C Metabolomics: NMR and IROA for unknown identification. *Curr. Metabolomics* **2016**, *4*, 116−120.

(201) Kazmi, S. R.; Jun, R.; Yu, M. S.; Jung, C.; Na, D. *Comput. Biol. Med.* **2019**, *106*, 54−64.

(202) Schuster, D.; Steindl, T.; Langer, T. Predicting drug metabolism induction in silico. *Curr. Top. Med. Chem.* **2006**, *6*, 1627−1640.

(203) Andrade, C.; Silva, D.; Braga, R. In silico prediction of drug metabolism by P450. *Curr. Drug Metab.* **2014**, *15*, 514−525.

(204) Willoughby, P. H.; Jansma, M. J.; Hoye, T. R. A guide to small-molecule structure assignment through computation of (1 H and 13 C) NMR chemical shifts. *Nat. Protoc.* **2014**, *9*, 643−660.

(205) Yu, Z.; Li, P.; Merz, K. M. Using ligand-induced protein chemical shift perturbations to determine protein-ligand structures. *Biochemistry* **2017**, *56*, 2349−2362.

(206) Wang, B.; Raha, K.; Merz, K. M., Jr Pose scoring by NMR. *J. Am. Chem. Soc.* **2004**, *126*, 11430−11431.

(207) Wang, B.; Merz, K. M., Jr Validation of the binding site structure of the cellular retinol-binding protein (CRBP) by ligand NMR chemical shift perturbations. *J. Am. Chem. Soc.* **2005**, *127*, 5310−5311.

(208) Wang, B.; Westerhoff, L. M.; Merz, K. M., Jr A critical assessment of the performance of protein-ligand scoring functions based on NMR chemical shift perturbations. *J. Med. Chem.* **2007**, *50*, 5128−5134.

(209) Yesiltepe, Y.; Nuñez, J. R.; Colby, S. M.; Thomas, D. G.; Borkum, M. I.; Reardon, P. N.; Washton, N. M.; Metz, T. O.; Teeguarden, J. G.; Govind, N.; Renslow, R. S. An automated framework for NMR chemical shift calculations of small organic molecules. *J. Cheminf.* **2018**, *10*, 52.

(210) Grimme, S.; Bannwarth, C.; Dohm, S.; Hansen, A.; Pisarek, J.; Pracht, P.; Seibert, J.; Neese, F. Fully automated quantum-chemistry-based computation of spin-spin-coupled nuclear magnetic resonance spectra. *Angew. Chem., Int. Ed.* **2017**, *56*, 14763−14769.

(211) Kaupp, M.; Bühl, M.; Malkin, V. G. *Calculation of NMR and EPR Parameters: Theory and Applications*; John Wiley & Sons, 2006.

(212) Casabianca, L. B.; De Dios, A. C. Ab initio calculations of NMR chemical shifts. *J. Chem. Phys.* **2008**, *128*, 052201.

(213) Bryce, D. L.; Wasylishen, R. E. Ab initio calculations of NMR parameters for diatomic molecules. an exercise in computational chemistry. *J. Chem. Educ.* **2001**, *78*, 124.

(214) Oldfield, E. Chemical shifts in amino acids, peptides, and proteins: from quantum chemistry to drug design. *Annu. Rev. Phys. Chem.* **2002**, *53*, 349−378.

(215) Facelli, J. C. Calculations of chemical shieldings: Theory and applications. *Concepts Magn. Reson.* **2004**, *20A*, 42−69.

(216) Auer, A. A.; Gauss, J.; Stanton, J. F. Quantitative prediction of gas-phase 13C nuclear magnetic shielding constants. *J. Chem. Phys.* **2003**, *118*, 10407−10417.

(217) Kalinichev, A. Molecular Simulations of Liquid and Supercritical Water: Thermodynamics, Structure, and Hydrogen Bonding. In *Molecular Modeling Theory: Applications in the Geosciences*; Cygan, R. T., Kubicki, J. D., Eds.; Mineralogical Society of America: Washington DC, 2001; *42*, pp 83−130.

(218) Sebastiani, D. Ab-initio calculations of NMR parameters in condensed phases. *Mod. Phys. Lett. B* **2003**, *17*, 1301−1319.

(219) Benzi, C.; Crescenzi, O.; Pavone, M.; Barone, V. Reliable NMR chemical shifts for molecules in solution by methods rooted in density functional theory. *Magn. Reson. Chem.* **2004**, *42*, S57−S67.

(220) Bagno, A.; Rastrelli, F.; Saielli, G. NMR techniques for the investigation of solvation phenomena and non-covalent interactions. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *47*, 41−93.

(221) Oldfield, E. Quantum chemical studies of protein structure. *Philos. Trans. R. Soc., B* **2005**, *360*, 1347−1361.

(222) Hunter, C. A.; Packer, M. J.; Zonta, C. From structure to chemical shift and vice-versa. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *47*, 27−39.

(223) Williams, D. E.; Peters, M. B.; Wang, B.; Roitberg, A. E.; Merz, K. M., Jr AMI parameters for the prediction of1H and13C NMR chemical shifts in proteins. *J. Phys. Chem. A* **2009**, *113*, 11550−11559.

(224) Wang, B.; Merz, K. M., Jr A fast QM/MM (quantum mechanical/molecular mechanical) approach to calculate nuclear magnetic resonance chemical shifts for macromolecules. *J. Chem. Theory Comput.* **2006**, *2*, 209−215.

(225) He, X.; Wang, B.; Merz, K. M., Jr Protein NMR chemical shift calculations based on the automated fragmentation QM/MM approach. *J. Phys. Chem. B* **2009**, *113*, 10380−10388.

(226) Bryce, D. L.; Sward, G. D. Solid-state NMR spectroscopy of the quadrupolar halogens: chlorine-35/37, bromine-79/81, and iodine-127. *Magn. Reson. Chem.* **2006**, *44*, 409−450.

(227) Saielli, G.; Nicolaou, K. C.; Ortiz, A.; Zhang, H.; Bagno, A. Addressing the stereochemistry of complex organic molecules by density functional theory-NMR: Vannusal B in retrospective. *J. Am. Chem. Soc.* **2011**, *133*, 6072−6077.

(228) Grimblat, N.; Zanardi, M. M.; Sarotti, A. M. Beyond DP4: an improved probability for the stereochemical assignment of isomeric compounds using quantum chemical calculations of NMR shifts. *J. Org. Chem.* **2015**, *80*, 12526−12534.

(229) Smith, S. G.; Goodman, J. M. Assigning stereochemistry to single diastereoisomers by GIAO NMR calculation: The DP4 probability. *J. Am. Chem. Soc.* **2010**, *132*, 12946−12959.

(230) Chimichi, S.; Boccalini, M.; Matteucci, A.; Kharlamov, S. V.; Latypov, S. K.; Sinyashin, O. G. GIAO DFT 13C/15N chemical shifts in regioisomeric structure determination of fused pyrazoles. *Magn. Reson. Chem.* **2010**, *48*, 607−613.

(231) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. Computational prediction of 1H and 13C chemical shifts: a useful tool for natural product, mechanistic, and synthetic organic chemistry. *Chem. Rev.* **2012**, *112*, 1839−1862.

(232) Lodewyk, M. W.; Tantillo, D. J. Prediction of the structure of nobilisitine A using computed NMR chemical shifts. *J. Nat. Prod.* **2011**, *74*, 1339−1343.

(233) Flaig, D.; Maurer, M.; Hanni, M.; Braunger, K.; Kick, L.; Thubauville, M.; Ochsenfeld, C. Benchmarking hydrogen and carbon NMR chemical shifts at HF, DFT, and MP2 Levels. *J. Chem. Theory Comput.* **2014**, *10*, 572−578.

(234) Teale, A. M.; Lutnæs, O. B.; Helgaker, T.; Tozer, D. J.; Gauss, J. Benchmarking density-functional theory calculations of NMR shielding constants and spin-rotation constants using accurate coupled-cluster calculations. *J. Chem. Phys.* **2013**, *138*, 024111.

(235) Pierens, G. K. 1H and 13C NMR scaling factors for the calculation of chemical shifts in commonly used solvents using density functional theory. *J. Comput. Chem.* **2014**, *35*, 1388−1394.

(236) Webb, G. A., Ed.; *Nuclear Magnetic Resonance*; Specialist Periodical Reports; Royal Society of Chemistry, London, 1985.

(237) Spivey, J. J. *Catalysis*; Royal Society of Chemistry, 2004; Vol. *17*.

(238) Merrill, A. T.; Tantillo, D. J. Solvent optimization and conformational flexibility effects on 1H and 13C NMR scaling factors. *Magn. Reson. Chem.* **2020**, *58*, 576−583.

(239) Xin, D.; Sader, C. A.; Chaudhary, O.; Jones, P. J.; Wagner, K.; Tautermann, C. S.; Yang, Z.; Busacca, C. A.; Saraceno, R. A.; Fandrick, K. R.; Gonnella, N. C.; Horspool, K.; Hansen, G.; Senanayake, C. H. Development of a 13C NMR chemical shift prediction procedure using B3LYP/cc-pVDZ and empirically derived systematic error correction terms: A computational small molecule structure elucidation method. *J. Org. Chem.* **2017**, *82*, 5135−5145.

(240) Merz, K. M., Jr Using quantum mechanical approaches to study biological systems. *Acc. Chem. Res.* **2014**, *47*, 2804−2811.

(241) Bühl, M.; Kaupp, M.; Malkina, O. L.; Malkin, V. G. The DFT route to NMR chemical shifts. *J. Comput. Chem.* **1999**, *20*, 91−105.

(242) Gertrudes, J. C.; Maltarollo, V. G.; Silva, R. A.; Oliveira, P. R.; Honorio, K. M.; da Silva, A. B. Machine learning techniques and drug design. *Curr. Med. Chem.* **2012**, *19*, 4289−4297.

(243) Rupp, M.; Ramakrishnan, R.; Von Lilienfeld, O. A. Machine learning for quantum mechanical properties of atoms in molecules. *J. Phys. Chem. Lett.* **2015**, *6*, 3309−3313.

(244) Das, S.; Edison, A. S.; Merz, K. M. Metabolite structure assignment using in silico NMR techniques. *Anal. Chem.* **2020**, *92*, 10412−10419.

(245) Lanucara, F.; Holman, S. W.; Gray, C. J.; Eyers, C. E. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. *Nat. Chem.* **2014**, *6*, 281−294.

(246) Paglia, G.; Astarita, G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.* **2017**, *12*, 797−813.

(247) Stow, S. M.; Causon, T. J.; Zheng, X. Y.; Kurulugama, R. T.; Mairinger, T.; May, J. C.; Rennie, E. E.; Baker, E. S.; Smith, R. D.; McLean, J. A.; Hann, S.; Fjeldsted, J. C. An interlaboratory evaluation of drift tube ion mobility-mass spectrometry collision cross section measurements. *Anal. Chem.* **2017**, *89*, 9048−9055.

(248) Nunez, J.; Thomas, D.; Colby, S.; Tfaily, M.; Tolic, N.; Metz, T.; Teeguarden, J.; Renslow, R. Standards-free identification of small molecules using multi-feature matching. *Abstr. Pap. Am. Chem. Soc.* **2019**, *257*.

(249) Shvartsburg, A. A.; Jarrold, M. F. An exact hard-spheres scattering model for the mobilities of polyatomic ions. *Chem. Phys. Lett.* **1996**, *261*, 86−91.

(250) Mesleh, M. F.; Hunter, J. M.; Shvartsburg, A. A.; Schatz, G. C.; Jarrold, M. F. Structural information from ion mobility measurements: Effects of the long-range potential. *J. Phys. Chem.* **1996**, *100*, 16082−16086.

(251) Gidden, J.; Bowers, M. T. Gas-phase conformations of deprotonated and protonated mononucleotides determined by ion mobility and theoretical modeling. *J. Phys. Chem. B* **2003**, *107*, 12829−12837.

(252) Leavell, M. D.; Gaucher, S. P.; Leary, J. A.; Taraszka, J. A.; Clemmer, D. E. Conformational studies of Zn-ligand-hexose diastereomers using ion mobility measurements and density functional theory calculations. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 284−293.

(253) Rusyniak, M. J.; Ibrahim, Y. M.; Wright, D. L.; Khanna, S. N.; El-Shall, M. S. Gas-phase ion mobilities and structures of benzene cluster cations (C6H6)n(+), n = 2−6. *J. Am. Chem. Soc.* **2003**, *125*, 12001−12013.

(254) Colby, S. M.; Nuñez, J. R.; Hodas, N. O.; Corley, C. D.; Renslow, R. R. Deep Learning to generate in silico chemical property libraries and candidate molecules for small molecule identification in complex samples. *Anal. Chem.* **2020**, *92*, 1720−1729.

(255) M?llerup, C. B.; Mardal, M.; Dalsgaard, P. W.; Linnet, K.; Barron, L. P. Prediction of collision cross section and retention time for broad scope screening in gradient reversed-phase liquid chromatography-ion mobility-high resolution accurate mass spectrometry. *J. Chromatogr., A* **2018**, *1542*, 82−88.

(256) Plante, P. L.; Francovic-Fontaine, E.; May, J. C.; McLean, J. A.; Baker, E. S.; Laviolette, F.; Marchand, M.; Corbeil, J. Predicting ion mobility collision cross-sections using a deep neural network: deepCCS. *Anal. Chem.* **2019**, *91*, 5191−5199.

(257) Zhou, Z. W.; Shen, X. T.; Tu, J.; Zhu, Z. J. Large-scale prediction of collision cross-section values for metabolites in ion mobility-mass spectrometry. *Anal. Chem.* **2016**, *88*, 11084−11091.

(258) Zhou, Z. W.; Tu, J.; Xiong, X.; Shen, X. T.; Zhu, Z. J. LipidCCS: prediction of collision cross-section values for lipids with high precision to support ion mobility-mass spectrometry-based lipidomics. *Anal. Chem.* **2017**, *89*, 9559−9566.

(259) Wyttenbach, T.; Witt, M.; Bowers, M. T. On the stability of amino acid zwitterions in the gas phase: The influence of derivatization, proton affinity, and alkali ion addition. *J. Am. Chem. Soc.* **2000**, *122*, 3458−3464.

(260) Rizzo, T. R.; Boyarkin, O. V. Cryogenic Methods for the Spectroscopy of Large, Biomolecular Ions. In *Gas-Phase IR Spectroscopy and Structure of Biological Molecules*; Topics in Current Chemistry; Rijs, A. M., Oomens, J., Eds.; Springer, 2015; pp 43−97.

(261) Kamrath, M. Z.; Rizzo, T. R. Combining ion mobility and cryogenic spectroscopy for structural and analytical studies of biomolecular ions. *Acc. Chem. Res.* **2018**, *51*, 1487−1495.

(262) Dear, G. J.; Munoz-Muriedas, J.; Beaumont, C.; Roberts, A.; Kirk, J.; Williams, J. P.; Campuzano, I. Sites of metabolic substitution: investigating metabolite structures utilising ion mobility and molecular modelling. *Rapid Commun. Mass Spectrom.* **2010**, *24*, 3157−3162.

(263) Campuzano, I.; Bush, M. F.; Robinson, C. V.; Beaumont, C.; Richardson, K.; Kim, H.; Kim, H. I. Structural characterization of drug-like compounds by ion mobility mass spectrometry: comparison of theoretical and experimentally derived nitrogen collision cross sections. *Anal. Chem.* **2012**, *84*, 1026−1033.

(264) Cuyckens, F.; Wassvik, C.; Mortishire-Smith, R. J.; Tresadern, G.; Campuzano, I.; Claereboudt, J. Product ion mobility as a promising tool for assignment of positional isomers of drug metabolites. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 3497−3503.

(265) Colby, S. M.; Thomas, D. G.; Nunez, J. R.; Baxter, D. J.; Glaesemann, K. R.; Brown, J. M.; Pirrung, M. A.; Govind, N.; Teeguarden, J. G.; Metz, T. O.; Renslow, R. S. ISiCLE: A Quantum chemistry pipeline for establishing in silico collision cross section libraries. *Anal. Chem.* **2019**, *91*, 4346−4356.

(266) Martens, J.; Berden, G.; van Outersterp, R. E.; Kluijtmans, L. A. J.; Engelke, U. F.; van Karnebeek, C. D. M.; Wevers, R. A.; Oomens, J. Molecular identification in metabolomics using infrared ion spectroscopy. *Sci. Rep.* **2017**, *7*, 3363.

(267) Clugston, M.; Flemming, R. *Advanced Chemistry*; Oxford University Press: Oxford, UK, 2008.

(268) Devlin, F. J.; Stephens, P. J.; Cheeseman, J. R.; Frisch, M. J. Ab initio prediction of vibrational absorption and circular dichroism spectra of chiral natural products using density functional theory: Camphor and Fenchone. *J. Phys. Chem. A* **1997**, *101*, 6322−6333.

(269) Martens, J.; van Outersterp, R. E.; Vreeken, R. J.; Cuyckens, F.; Coene, K. L. M.; Engelke, U. F.; Kluijtmans, L. A. J.; Wevers, R. A.; Buydens, L. M. C.; Redlich, B.; Berden, G.; Oomens, J. Infrared ion

spectroscopy: New opportunities for small-molecule identification in mass spectrometry - A tutorial perspective. *Anal. Chim. Acta* **2020**, *1093*, 1−15.

(270) Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; et al. METLIN: A technology platform for identifying knowns and unknowns. *Anal. Chem.* **2018**, *90*, 3156−3164.

(271) Aseev, O.; Perez, M. A. S.; Rothlisberger, U.; Rizzo, T. R. Cryogenic spectroscopy and quantum molecular dynamics determine the structure of cyclic intermediates involved in peptide sequence scrambling. *J. Phys. Chem. Lett.* **2015**, *6*, 2524−2529.

(272) Ben Faleh, A.; Warnke, S.; Rizzo, T. R. Combining ultrahigh-resolution ion-mobility spectrometry with cryogenic infrared spectroscopy for the analysis of glycan mixtures. *Anal. Chem.* **2019**, *91*, 4876−4882.

(273) Khanal, N.; Masellis, C.; Kamrath, M. Z.; Clemmer, D. E.; Rizzo, T. R. Cryogenic IR spectroscopy combined with ion mobility spectrometry for the analysis of human milk oligosaccharides. *Analyst* **2018**, *143*, 1846−1852.

(274) Khanal, N.; Masellis, C.; Kamrath, M. Z.; Clemmer, D. E.; Rizzo, T. R. Glycosaminoglycan analysis by cryogenic messenger-tagging IR spectroscopy combined with IMS-MS. *Anal. Chem.* **2017**, *89*, 7601−7606.

(275) Scutelnic, V.; Rizzo, T. R. Cryogenic ion spectroscopy for identification of monosaccharide anomers. *J. Phys. Chem. A* **2019**, *123*, 2815−2819.

(276) Warnke, S.; Ben Faleh, A.; Pellegrinelli, R. P.; Yalovenko, N.; Rizzo, T. R. Combining ultra-high resolution ion mobility spectrometry with cryogenic IR spectroscopy for the study of biomolecular ions. *Faraday Discuss.* **2019**, *217*, 114−125.

(277) Bell, M. R.; Tesler, L. E.; Polfer, N. C. Cryogenic infrared ion spectroscopy for the structural elucidation of drug molecules: MDMA and its metabolites. *Int. J. Mass Spectrom.* **2019**, *443*, 101−108.

(278) Cismesia, A. P.; Bailey, L. S.; Bell, M. R.; Tesler, L. F.; Polfer, N. C. Making mass spectrometry see the light: The promises and challenges of cryogenic infrared ion spectroscopy as a bioanalytical technique. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 757−766.

(279) Cismesia, A. P.; Bell, M. R.; Tesler, L. F.; Alves, M.; Polfer, N. C. Infrared ion spectroscopy: an analytical tool for the study of metabolites. *Analyst* **2018**, *143*, 1615−1623.

(280) Cismesia, A. P.; Tesler, L. F.; Bell, M. R.; Bailey, L. S.; Polfer, N. C. Infrared ion spectroscopy inside a mass-selective cryogenic 2D linear ion trap. *J. Mass Spectrom.* **2017**, *52*, 720−727.

(281) Tesler, L. F.; Cismesia, A. P.; Bell, M. R.; Bailey, L. S.; Polfer, N. C. Operation and performance of a mass-selective cryogenic linear ion trap. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 2115−2124.

(282) Wassermann, T. N.; Boyarkin, O. V.; Paizs, B.; Rizzo, T. R. Conformation-specific spectroscopy of peptide fragment ions in a low-temperature ion trap. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1029−1045.

(283) Elyashberg, M.; Blinov, K.; Martirosian, E. A new approach to computer-aided molecular structure elucidation: the expert system Structure Elucidator. *Lab. Autom. Inf. Manage.* **1999**, *34*, 15−30.

(284) Wolfender, J.-L.; Waridel, P.; Ndjoko, K.; Hobby, K.; Major, H.; Hostettmann, K. Evaluation of Q-TOF-MS/MS and multiple stage IT-MSn for the dereplication of flavonoids and related compounds in crude plant extracts. *Analusis* **2000**, *28*, 895−906.

(285) Vaniya, A.; Fiehn, O. Using fragmentation trees and mass spectral trees for identifying unknown compounds in metabolomics. *TrAC, Trends Anal. Chem.* **2015**, *69*, 52−61.

(286) Ridder, L.; van der Hooft, J. J.; Verhoeven, S.; de Vos, R. C.; van Schaik, R.; Vervoort, J. Substructure-based annotation of high-resolution multistage MSn spectral trees. *Rapid Commun. Mass Spectrom.* **2012**, *26*, 2461−2471.

(287) Sheldon, M. T.; Mistrik, R.; Croley, T. R. Determination of ion structures in structurally related compounds using precursor ion fingerprinting. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 370−376.

(288) Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites* **2018**, *8*, 31.

(289) Schüler, J. A.; Neumann, S.; Müller-Hannemann, M.; Brandt, W. ChemFrag: chemically meaningful annotation of fragment ion mass spectra. *J. Mass Spectrom.* **2018**, *53*, 1104−1115.

(290) Grimme, S. Towards first principles calculation of electron impact mass spectra of molecules. *Angew. Chem., Int. Ed.* **2013**, *52*, 6306−6312.

(291) Peisl, L.; Schymanski, E. L.; Wilmes, P. Dark matter in host-microbiome metabolomics: tackling the unknowns-a review. *Anal. Chim. Acta* **2018**, *1037*, 13−27.

(292) Scott, D. 1-Aminopropane, 2-aminopropane, and 2-methyl-2-aminopropane: vibrational assignments, conformational analyses, and chemical thermodynamic properties. *J. Chem. Thermodyn.* **1971**, *3*, 843−852.

(293) Gross, J. H. *Mass Spectrometry: A Textbook*; Springer: Berlin, Heidelberg, 2011.

(294) Stein, S. Mass spectral reference libraries: an ever-expanding resource for chemical identification. *Anal. Chem.* **2012**, *84*, 7274−7282.

(295) Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S. S.; Wohlgemuth, G.; Barupal, D. K.; Showalter, M. R.; Arita, M.; Fiehn, O. Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom. Rev.* **2018**, *37*, 513−532.

(296) Armentrout, P. B.; Ervin, K. M.; Rodgers, M. T. Statistical rate theory and kinetic energy-resolved ion chemistry: Theory and applications. *J. Phys. Chem. A* **2008**, *112*, 10071−10085.

(297) Tureček, F.; Julian, R. R. Peptide radicals and cation radicals in the gas phase. *Chem. Rev.* **2013**, *113*, 6691−6733.

(298) Rennie, E. E.; Cooper, L.; Shpinkova, L. G.; Holland, D. M. P.; Shaw, D. A.; Guest, M. F.; Mayer, P. M. Methyl t-Butyl Ether and Methyl Trimethylsilyl Ether ions dissociate near their ionization thresholds: A TPES, TPEPICO, RRKM, and G3 investigation. *J. Phys. Chem. A* **2009**, *113*, 5823−5831.

(299) Tsyshevsky, R. V.; Garifzianova, G. G.; Shamov, A. G.; Khrapkovskii, G. M. Fragmentation reactions in the 1-nitropropane radical cation induced by γ-hydrogen shift: Ab initio study. *Int. J. Mass Spectrom.* **2014**, *369*, 36−43.

(300) Solano, E. A.; Mayer, P. M. A complete map of the ion chemistry of the naphthalene radical cation? DFT and RRKM modeling of a complex potential energy surface. *J. Chem. Phys.* **2015**, *143*, 104305.

(301) Rosenstock, H. M.; Wallenstein, M. B.; Wahrhaftig, A. L.; Eyring, H. Absolute rate theory for isolated systems and the mass spectra of polyatomic molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1952**, *38*, 667−678.

(302) Rice, O. K.; Ramsperger, H. C. Theories of unimolecular gas reactions at low pressures. *J. Am. Chem. Soc.* **1927**, *49*, 1617−1629.

(303) Kassel, L. S. Studies in homogeneous gas reactions. I. *J. Phys. Chem.* **1928**, *32*, 225−242.

(304) Marcus, R. A.; Rice, O. K. The kinetics of the recombination of Methyl radicals and Iodine atoms. *J. Phys. Chem.* **1951**, *55*, 894−908.

(305) Marcus, R. A. Unimolecular dissociations and free radical recombination reactions. *J. Chem. Phys.* **1952**, *20*, 359−364.

(306) Bauer, C. A.; Grimme, S. How to Compute electron ionization mass spectra from first principles. *J. Phys. Chem. A* **2016**, *120*, 3755−3766.

(307) Maeda, S.; Harabuchi, Y.; Ono, Y.; Taketsugu, T.; Morokuma, K. Intrinsic reaction coordinate: Calculation, bifurcation, and automated search. *Int. J. Quantum Chem.* **2015**, *115*, 258−269.

(308) Vázquez, S. A.; Otero, X. L.; Martinez-Nunez, E. A trajectory-based method to explore reaction mechanisms. *Molecules* **2018**, *23*, 3156.

(309) Döntgen, M.; Przybylski-Freund, M.-D.; Kröger, L. C.; Kopp, W. A.; Ismail, A. E.; Leonhard, K. Automated discovery of reaction pathways, rate constants, and transition states using reactive molecular dynamics simulations. *J. Chem. Theory Comput.* **2015**, *11*, 2517−2524.

(310) Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.* **2000**, *103*, 495−506.

(311) Cui, Q.; Elstner, M. Density functional tight binding: values of semi-empirical methods in an ab initio era. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14368−14377.

(312) Grimme, S.; Bannwarth, C.; Shushkov, P. A Robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for All spd-block elements (Z = 1−86). *J. Chem. Theory Comput.* **2017**, *13*, 1989−2009.

(313) Neese, F. Software update: the ORCA program system, version 4.0. *WIRES: Comput. Mol. Sci.* **2018**, *8*, No. e1327.

(314) Neese, F. The ORCA program system. *WIRES: Comput. Mol. Sci.* **2012**, *2*, 73−78.

(315) Furche, F.; Ahlrichs, R.; Hättig, C.; Klopper, W.; Sierka, M.; Weigend, F. Turbomole. *WIRES: Comput. Mol. Sci.* **2014**, *4*, 91−100.

(316) Stewart, J. J. MOPAC: a semiempirical molecular orbital program. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1−103.

(317) Dral, P. O.; Wu, X.; Thiel, W. Semiempirical quantum-chemical methods with Orthogonalization and dispersion corrections. *J. Chem. Theory Comput.* **2019**, *15*, 1743−1760.

(318) Dewar, M. J. S.; Thiel, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.

(319) Ásgeirsson, V.; Bauer, C. A.; Grimme, S. Quantum chemical calculation of electron ionization mass spectra for general organic and inorganic molecules. *Chemical Science* **2017**, *8*, 4879−4895.

(320) Koopman, J.; Grimme, S. Calculation of Electron Ionization Mass Spectra with Semiempirical GFNn-xTB Methods. *ACS Omega* **2019**, *4*, 15120−15133.

(321) Zaikin, V.; Halket, J. M. *A Handbook of Derivatives for Mass Spectrometry*; IM Publications, 2009.

(322) Bauer, C. A.; Grimme, S. Automated quantum chemistry based molecular dynamics simulations of electron ionization induced fragmentations of the nucleobases Uracil, Thymine, Cytosine, and Guanine. *Eur. J. Mass Spectrom.* **2015**, *21*, 125−140.

(323) Cautereels, J.; Claeys, M.; Geldof, D.; Blockhuys, F. Quantum chemical mass spectrometry: ab initio prediction of electron ionization mass spectra and identification of new fragmentation pathways. *J. Mass Spectrom.* **2016**, *51*, 602−614.

(324) Morton, T. H. Neutral products from gas phase rearrangements of simple carbocations. *Adv. Gas Phase Ion Chem.* **2001**, *4*, 213−256.

(325) Allen, F.; Pon, A.; Greiner, R.; Wishart, D. Computational prediction of electron ionization mass spectra to assist in GC/MS compound identification. *Anal. Chem.* **2016**, *88*, 7689−7697.

(326) Spackman, P. R.; Bohman, B.; Karton, A.; Jayatilaka, D. Quantum chemical electron impact mass spectrum prediction for de novo structure elucidation: Assessment against experimental reference data and comparison to competitive fragmentation modeling. *Int. J. Quantum Chem.* **2018**, *118*, No. e25460.

(327) Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Thiel, W. Semiempirical quantum-chemical orthogonalization-corrected methods: Benchmarks for ground-state properties. *J. Chem. Theory Comput.* **2016**, *12*, 1097−1120.

(328) Stewart, J. J. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1−32.

(329) Bauer, C. A.; Grimme, S. Elucidation of electron ionization induced fragmentations of Adenine by semiempirical and density functional molecular dynamics. *J. Phys. Chem. A* **2014**, *118*, 11479−11484.

(330) Giorgi, G. In *Detection of Chemical, Biological, Radiological and Nuclear Agents for the Prevention of Terrorism*; Springer, 2014; pp 17−31.

(331) Kussmann, J. R.; Ochsenfeld, C. Hybrid CPU/GPU integral engine for strong-scaling ab initio methods. *J. Chem. Theory Comput.* **2017**, *13*, 3153−3159.

(332) Bauer, C. A.; Hansen, A.; Grimme, S. The fractional occupation number weighted density as a versatile analysis tool for molecules with a complicated electronic structure. *Chem. - Eur. J.* **2017**, *23*, 6150−6164.

(333) Keshet, U.; Goldshlag, P.; Amirav, A. Pesticide analysis by pulsed flow modulation GCxGC-MS with Cold EI—an alternative to GC-MS-MS. *Anal. Bioanal. Chem.* **2018**, *410*, 5507−5519.

(334) Buchalter, S.; Marginean, I.; Yohannan, J.; Lurie, I. S. Gas chromatography with tandem cold electron ionization mass spectrometric detection and vacuum ultraviolet detection for the comprehensive analysis of fentanyl analogues. *Journal of Chromatography A* **2019**, *1596*, 183−193.

(335) Jacovella, U.; da Silva, G.; Bieske, E. J. Unveiling new isomers and rearrangement routes on the C7H8+ potential energy surface. *J. Phys. Chem. A* **2019**, *123*, 823−830.

(336) Wu, X.; Zhou, X.; Hemberger, P.; Bodi, A. A guinea pig for conformer selectivity and mechanistic insights into dissociative ionization by photoelectron photoion coincidence: fluorocyclohexane. *Phys. Chem. Chem. Phys.* **2020**, *22*, 2351−2360.

(337) Candian, A.; Bouwman, J.; Hemberger, P.; Bodi, A.; Tielens, A. G. Dissociative ionisation of adamantane: a combined theoretical and experimental study. *Phys. Chem. Chem. Phys.* **2018**, *20*, 5399−5406.

(338) Majer, K.; Signorell, R.; Heringa, M. F.; Goldmann, M.; Hemberger, P.; Bodi, A. Valence photoionization of Thymine: ionization energies, vibrational structure, and fragmentation pathways from the slow to the ultrafast. *Chem. - Eur. J.* **2019**, *25*, 14192−14204.

(339) Johnson, J. V.; Yost, R. A.; Kelley, P. E.; Bradford, D. C. Tandem-in-space and tandem-in-time mass spectrometry: triple quadrupoles and quadrupole ion traps. *Anal. Chem.* **1990**, *62*, 2162−2172.

(340) McLafferty, F.; Bryce, T. Metastable-ion characteristics: characterization of isomeric molecules. *Chem. Commun. (London)* **1967**, 1215−1217.

(341) Jennings, K. R. Collision-induced decompositions of aromatic molecular ions. *Int. J. Mass Spectrom. Ion Phys.* **1968**, *1*, 227−235.

(342) Yost, R.; Enke, C. Selected ion fragmentation with a tandem quadrupole mass spectrometer. *J. Am. Chem. Soc.* **1978**, *100*, 2274−2275.

(343) Yost, R.; Enke, C. Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal. Chem.* **1979**, *51*, 1251−1264.

(344) Douglas, D. Mechanism of the collision-induced dissociation of polyatomic ions studied by triple quadrupole mass spectrometry. *J. Phys. Chem.* **1982**, *86*, 185−191.

(345) Dawson, P.; Fulford, J. The effective containment of parent ions and daughter ions in triple quadrupoles used for collisional dissociation. *Int. J. Mass Spectrom. Ion Phys.* **1982**, *42*, 195−211.

(346) Dawson, P. A study of the collision-induced dissociation of C2H5OH2+ using various target gases. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *50*, 287−297.

(347) Vinaixa, M.; Schymanski, E. L.; Neumann, S.; Navarro, M.; Salek, R. M.; Yanes, O. Mass spectral databases for LC/MS-and GC/MS-based metabolomics: state of the field and future prospects. *TrAC, Trends Anal. Chem.* **2016**, *78*, 23−35.

(348) Niessen, W. M.; Correa, C. R. A. *Interpretation of MS−MS mass spectra of drugs and pesticides*; John Wiley & Sons, 2017.

(349) Niessen, W. MSMS and MSn; *Reference Module in Chemical, Molecular Sciences and Engineering*; Elsevier, 2013.

(350) Vékey, K. Internal energy effects in mass spectrometry. *J. Mass Spectrom.* **1996**, *31*, 445−463.

(351) Demarque, D. P.; Crotti, A. E.; Vessecchi, R.; Lopes, J. L.; Lopes, N. P. Fragmentation reactions using electrospray ionization mass spectrometry: an important tool for the structural elucidation and characterization of synthetic and natural products. *Nat. Prod. Rep.* **2016**, *33*, 432−455.

(352) Mayer, P. M.; Poon, C. The mechanisms of collisional activation of ions in mass spectrometry. *Mass Spectrom. Rev.* **2009**, *28*, 608−639.

(353) Spezia, R.; Martin-Somer, A.; Macaluso, V.; Homayoon, Z.; Pratihar, S.; Hase, W. L. Unimolecular dissociation of peptides: statistical vs. non-statistical fragmentation mechanisms and time scales. *Faraday Discuss.* **2016**, *195*, 599−618.

(354) Martínez-Núñez, E.; Fernández-Ramos, A.; Vázquez, S. A.; Marques, J. M.; Xue, M.; Hase, W. L. Quasiclassical dynamics simulation of the collision-induced dissociation of Cr (CO) 6+ with Xe. *J. Chem. Phys.* **2005**, *123*, 154311.

(355) Martínez-Núñez, E.; Vázquez, S. A.; Marques, J. Quasiclassical trajectory study of the collision-induced dissociation of CH3SH++ Ar. *J. Chem. Phys.* **2004**, *121*, 2571−2577.

(356) Linhananta, A.; Lim, K. F. Quasiclassical trajectory calculations of collisional energy transfer in propane systems. *Phys. Chem. Chem. Phys.* **2000**, *2*, 1385−1392.

(357) Galezowska, A.; Harrison, M. W.; Herniman, J. M.; Skylaris, C. K.; Langley, G. J. A predictive science approach to aid understanding of electrospray ionisation tandem mass spectrometric fragmentation pathways of small molecules using density functional calculations. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 964−970.

(358) Alex, A.; Harvey, S.; Parsons, T.; Pullen, F. S.; Wright, P.; Riley, J. A. Can density functional theory (DFT) be used as an aid to a deeper understanding of tandem mass spectrometric fragmentation pathways? *Rapid Commun. Mass Spectrom.* **2009**, *23*, 2619−2627.

(359) Wright, P.; Alex, A.; Harvey, S.; Parsons, T.; Pullen, F. Understanding collision-induced dissociation of dofetilide: a case study in the application of density functional theory as an aid to mass spectral interpretation. *Analyst* **2013**, *138*, 6869−6880.

(360) Wright, P.; Alex, A.; Pullen, F. Predicting collision-induced dissociation spectra: Semi-empirical calculations as a rapid and effective tool in software-aided mass spectral interpretation. *Rapid Commun. Mass Spectrom.* **2014**, *28*, 1127−1143.

(361) Janesko, B. G.; Li, L.; Mensing, R. Quantum Chemical Fragment Precursor Tests: Accelerating de novo annotation of tandem mass spectra. *Anal. Chim. Acta* **2017**, *995*, 52−64.

(362) Gu, M.; Zhang, J.; Hase, W. L.; Yang, L. Direct Dynamics Simulations of the Thermal Fragmentation of a Protonated Peptide Containing Arginine. *ACS Omega* **2020**, *5*, 1463−1471.

(363) Lourderaj, U.; Sun, R.; Kohale, S. C.; Barnes, G. L.; de Jong, W. A.; Windus, T. L.; Hase, W. L. The VENUS/NWChem software package. Tight coupling between chemical dynamics simulations and electronic structure theory. *Comput. Phys. Commun.* **2014**, *185*, 1074−1080.

(364) Jara-Toro, R. A.; Pino, G. A.; Glowacki, D. R.; Shannon, R. J.; Martínez-Núñez, E. Enhancing automated reaction discovery with boxed molecular dynamics in energy space. *ChemSystemsChem* **2020**, *2*, No. e1900024.

(365) Macaluso, V.; Scuderi, D.; Crestoni, M. E.; Fornarini, S.; Corinti, D.; Dalloz, E.; Martinez-Nunez, E.; Hase, W. L.; Spezia, R. L-Cysteine modified by S-sulfation: consequence on fragmentation processes elucidated by tandem mass spectrometry and chemical dynamics simulations. *J. Phys. Chem. A* **2019**, *123*, 3685−3696.

(366) Lee, G.; Park, E.; Chung, H.; Jeanvoine, Y.; Song, K.; Spezia, R. Gas phase fragmentation mechanisms of protonated testosterone as revealed by chemical dynamics simulations. *Int. J. Mass Spectrom.* **2016**, *407*, 40−50.

(367) Krishnan, Y.; Sharma, N.; Lourderaj, U.; Paranjothy, M. Classical dynamics simulations of dissociation of protonated Tryptophan in the gas phase. *J. Phys. Chem. A* **2017**, *121*, 4389−4396.

(368) Spezia, R.; Salpin, J.-Y.; Gaigeot, M.-P.; Hase, W. L.; Song, K. Protonated urea collision-induced dissociation. Comparison of experiments and chemical dynamics simulations. *J. Phys. Chem. A* **2009**, *113*, 13853−13862.

(369) Spezia, R.; Martens, J.; Oomens, J.; Song, K. Collision-induced dissociation pathways of protonated Gly2NH2 and Gly3NH2 in the short time-scale limit by chemical dynamics and ion spectroscopy. *Int. J. Mass Spectrom.* **2015**, *388*, 40−52.

(370) Meroueh, S. O.; Wang, Y.; Hase, W. L. Direct dynamics simulations of collision- and surface-induced dissociation of N-Protonated Glycine. shattering fragmentation. *J. Phys. Chem. A* **2002**, *106*, 9983−9992.

(371) Park, K.; Deb, B.; Song, K.; Hase, W. L. Importance of shattering fragmentation in the surface-induced dissociation of protonated octaglycine. *J. Am. Soc. Mass Spectrom.* **2009**, *20*, 939−948.

(372) Ortiz, D.; Salpin, J.-Y.; Song, K.; Spezia, R. Galactose-6-Sulfate collision induced dissociation using QM+ MM chemical dynamics simulations and ESI-MS/MS experiments. *Int. J. Mass Spectrom.* **2014**, *358*, 25−35.

(373) Song, K.; Spezia, R. *Theoretical Mass Spectrometry: Tracing Ions with Classical Trajectories*; Walter de Gruyter, 2018.

(374) Cautereels, J.; Blockhuys, F. Quantum chemical mass spectrometry: verification and extension of the mobile proton model for histidine. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1227−1235.

(375) Bythell, B. J. Comment on: "Quantum chemical mass spectrometry: Verification and extension of the mobile proton model for Histidine" by Julie Cautereels and Frank Blockhuys, J. Am. Soc. Mass Spectrom. 28, 1227−1235 (2017). *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 2728−2730.

(376) Cautereels, J.; Van Hee, N.; Chatterjee, S.; Van Alsenoy, C.; Lemière, F.; Blockhuys, F. QCMS2 as a new method for providing insight into peptide fragmentation: The influence of the side-chain and inter-side-chain interactions. *J. Mass Spectrom.* **2019**, *55*, e4446.

(377) Carrà, A.; Macaluso, V.; Villalta, P. W.; Spezia, R.; Balbo, S. Fragmentation spectra prediction and DNA adducts structural determination. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2771−2784.

(378) Martin Somer, A.; Macaluso, V.; Barnes, G. L.; Yang, L.; Pratihar, S.; Song, K.; Hase, W. L.; Spezia, R. Role of chemical dynamics simulations in mass spectrometry studies of collision-induced dissociation and collisions of biological ions with organic surfaces. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 2−24.

(379) Pracht, P.; Bauer, C. A.; Grimme, S. Automated and efficient quantum chemical determination and energetic ranking of molecular protonation sites. *J. Comput. Chem.* **2017**, *38*, 2618−2631.

(380) Huang, N.; Siegel, M. M.; Kruppa, G. H.; Laukien, F. H. Automation of a fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-resolution exact-mass electrospray ionization mass spectral data. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 1166−1173.

(381) Zhu, J.; Cole, R. B. Formation and decompositions of chloride adduct ions,[M+ Cl]-, in negative ion electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 932−941.

(382) Tautenhahn, R.; Böttcher, C.; Neumann, S. Annotation of LC/ESI-MS Mass Signals. In *Bioinformatics Research and Development; First International Conference, BIRD 2007, Berlin, Germany, March 12-14, 2007, Proceedings*; Springer, 2007; pp 371−380.

(383) Ciminiello, P.; Dell'Aversano, C.; Dello Iacovo, E.; Fattorusso, E.; Forino, M.; Grauso, L.; Tartaglione, L. High resolution LC-MSn fragmentation pattern of palytoxin as template to gain new insights into ovatoxin-a structure. The key role of calcium in MS behavior of palytoxins. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 952−963.

(384) Trujillo, C.; Lamsabhi, A. M.; Mó, O.; Yáñez, M.; Salpin, J.-Y. Unimolecular reactivity upon collision of uracil-Ca2+ complexes in the gas phase: Comparison with uracil-M+ (M= H, alkali metals) and uracil-M2+ (M= Cu, Pb) systems. *Int. J. Mass Spectrom.* **2011**, *306*, 27−36.

(385) Chiu, C.-C.; Tsai, S.-T.; Hsu, P.-J.; Huynh, H. T.; Chen, J.-L.; Phan, H. T.; Huang, S.-P.; Lin, H.-Y.; Kuo, J.-L.; Ni, C.-K. Unexpected dissociation mechanism of Sodiated N-Acetylglucosamine and N-Acetylgalactosamine. *J. Phys. Chem. A* **2019**, *123*, 3441−3453.

(386) Homayoon, Z.; Macaluso, V.; Martin-Somer, A.; Muniz, M. C. N. B.; Borges, I.; Hase, W. L.; Spezia, R. Chemical dynamics simulations of CID of peptide ions: comparisons between TIK (H+) 2 and TLK (H+) 2 fragmentation dynamics, and with thermal simulations. *Phys. Chem. Chem. Phys.* **2018**, *20*, 3614−3629.

(387) Aribi, H. E.; Rodriquez, C. F.; Almeida, D. R.; Ling, Y.; Mak, W. W.-N.; Hopkinson, A. C.; Siu, K. M. Elucidation of fragmentation mechanisms of protonated peptide ions and their products: a case

study on glycylglycylglycine using density functional theory and threshold collision-induced dissociation. *J. Am. Chem. Soc.* **2003**, *125*, 9229−9236.

(388) Rossich Molina, E. R.; Eizaguirre, A.; Haldys, V.; Urban, D.; Doisneau, G.; Bourdreux, Y.; Beau, J. M.; Salpin, J. Y.; Spezia, R. Characterization of protonated model disaccharides from tandem mass spectrometry and chemical dynamics simulations. *ChemPhysChem* **2017**, *18*, 2812−2823.

(389) Wolken, J. K.; Tureček, F. Proton affinity of uracil. A computational study of protonation sites. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 1065−1071.

(390) Moser, A.; Range, K.; York, D. M. Accurate proton affinity and gas-phase basicity values for molecules important in biocatalysis. *J. Phys. Chem. B* **2010**, *114*, 13911−13921.

(391) Wright, P.; Alex, A.; Pullen, F. Predicting collision-induced dissociation mass spectra: understanding the role of the mobile proton in small molecule fragmentation. *Rapid Commun. Mass Spectrom.* **2016**, *30*, 1163−1175.

(392) Attygalle, A. B.; Xia, H.; Pavlov, J. Influence of ionization source conditions on the gas-phase protomer distribution of anilinium and related cations. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1575−1586.

(393) Grimme, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J. Chem. Theory Comput.* **2019**, *15*, 2847−2862.

(394) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, *15*, 1652−1671.

(395) Lapthorn, C.; Dines, T. J.; Chowdhry, B. Z.; Perkins, G. L.; Pullen, F. S. Can ion mobility mass spectrometry and density functional theory help elucidate protonation sites in'small'molecules? *Rapid Commun. Mass Spectrom.* **2013**, *27*, 2399−2410.

(396) Lu, W.; Liu, J. Deprotonated guanine· cytosine and 9-methylguanine· cytosine base pairs and their "non-statistical" kinetics: a combined guided-ion beam and computational study. *Phys. Chem. Chem. Phys.* **2016**, *18*, 32222−32237.

(397) Zhachkina, A.; Liu, M.; Sun, X.; Amegayibor, F. S.; Lee, J. K. Gas-phase thermochemical properties of the damaged base O 6-methylguanine versus adenine and guanine. *J. Org. Chem.* **2009**, *74*, 7429−7440.

(398) Araya-Sibaja, A. M.; Urgellés, M.; Vásquez-Castro, F.; Vargas-Huertas, F.; Vega-Baudrit, J. R.; Guillén-Girón, T.; Navarro-Hoyos, M.; Cuffini, S. L. The effect of solution environment and the electrostatic factor on the crystallisation of desmotropes of irbesartan. *RSC Adv.* **2019**, *9*, 5244−5250.

(399) Furlong, J.; Schiavoni, M. M.; Castro, E.; Allegretti, P. Mass spectrometry as a tool for studying tautomerism. *Russ. J. Org. Chem.* **2008**, *44*, 1725−1736.

(400) Terent'ev, P.; Kalandarishvili, A. Application of mass spectrometry for the analysis of organic tautomeric compounds. *Mass Spectrom. Rev.* **1996**, *15*, 339−363.

(401) Sayle, R. A. So you think you understand tautomerism? *J. Comput.-Aided Mol. Des.* **2010**, *24*, 485−496.

(402) Watson, M. A.; Yu, H. S.; Bochevarov, A. D. Generation of Tautomers using Micro-pKa's. *J. Chem. Inf. Model.* **2019**, *59*, 2672−2689.

(403) Kochev, N. T.; Paskaleva, V. H.; Jeliazkova, N. Ambit-Tautomer: an open source tool for Tautomer generation. *Mol. Inf.* **2013**, *32*, 481−504.

(404) Spjuth, O.; Berg, A.; Adams, S.; Willighagen, E. L. Applications of the InChI in cheminformatics with the CDK and Bioclipse. *J. Cheminf.* **2013**, *5*, 14.

(405) Landrum, G. *RDKit: Open-Source Cheminformatics Software*, 2006.

(406) Will, T.; Hutter, M. C.; Jauch, J.; Helms, V. Batch tautomer generation with MolTPC. *J. Comput. Chem.* **2013**, *34*, 2485−2492.

(407) Lu, H.-J.; Guo, Y.-L. Evaluation of chiral recognition characteristics of metal and proton complexes of di-o-benzoyl-tartaric acid dibutyl ester and L-tryptophan in the gas phase. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 571−580.

(408) Colby, S. M.; Thomas, D. G.; Nunez, J. R.; Baxter, D. J.; Glaesemann, K. R.; Brown, J. M.; Pirrung, M. A.; Govind, N.; Teeguarden, J. G.; Metz, T. O.; Renslow, R. S. ISiCLE: A quantum chemistry pipeline for establishing in silico collision cross section libraries. *Anal. Chem.* **2019**, *91*, 4346−4356.

(409) Gelpí, E. From large analogical instruments to small digital black boxes: 40 years of progress in mass spectrometry and its role in proteomics. Part II 1985−2000. *J. Mass Spectrom.* **2009**, *44*, 1137−1161.

(410) Oberacher, H.; Reinstadler, V.; Kreidl, M.; Stravs, M.; Hollender, J.; Schymanski, E. Annotating nontargeted LC-HRMS/MS data with two complementary tandem mass spectral libraries. *Metabolites* **2019**, *9*, 3.

(411) Holmes, J. L. Assigning structures to ions in the gas phase. *Org. Mass Spectrom.* **1985**, *20*, 169−183.

(412) Zhou, M.; Huang, C.; Wysocki, V. H. Surface-induced dissociation of ion mobility-separated noncovalent complexes in a quadrupole/time-of-flight mass spectrometer. *Anal. Chem.* **2012**, *84*, 6016−6023.

(413) Champarnaud, E.; Hopley, C. Evaluation of the comparability of spectra generated using a tuning point protocol on twelve electrospray ionisation tandem-in-space mass spectrometers. *Rapid Commun. Mass Spectrom.* **2011**, *25*, 1001−1007.

(414) Bristow, A. W.; Webb, K. S.; Lubben, A. T.; Halket, J. Reproducible product-ion tandem mass spectra on various liquid chromatography/mass spectrometry instruments for the development of spectral libraries. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1447−1454.

(415) Kind, T.; Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. *Bioanalytical reviews* **2010**, *2*, 23−60.

(416) Kandala, A.; Mezzacapo, A.; Temme, K.; Takita, M.; Brink, M.; Chow, J. M.; Gambetta, J. M. Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **2017**, *549*, 242−246.

(417) Colless, J. I.; Ramasesh, V. V.; Dahlen, D.; Blok, M. S.; Kimchi-Schwartz, M.; McClean, J.; Carter, J.; De Jong, W.; Siddiqi, I. Computation of molecular spectra on a quantum processor with an error-resilient algorithm. *Phys. Rev. X* **2018**, *8*, 011021.

(418) Thackston, R.; Fortenberry, R. C. The performance of low-cost commercial cloud computing as an alternative in computational chemistry. *J. Comput. Chem.* **2015**, *36*, 926−933.

(419) Kurtzer, G. M.; Sochat, V.; Bauer, M. W. Singularity: Scientific containers for mobility of compute. *PLoS One* **2017**, *12*, No. e0177459.

(420) Arango, C.; Dernat, R.; Sanabria, J. Performance evaluation of container-based virtualization for high performance computing environments; *arXiv* **2017**.1709.10140v1.

(421) Goerigk, L.; Mehta, N. A Trip to the density functional theory zoo: Warnings and recommendations for the user. *Aust. J. Chem.* **2019**, *72*, 563−573.

(422) Tantillo, D. J.*Applied Theoretical Organic Chemistry*; World Scientific Publishing, 2017.

(423) Bergmann, T. G.; Welzel, M. O.; Jacob, C. R. Towards theoretical spectroscopy with error bars: systematic quantification of the structural sensitivity of calculated spectra. *Chemical Science* **2020**, *11*, 1862−1877.

(424) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16*; Gaussian, Inc.: Wallingford, CT, 2016.

(425) Roothaan, C. C. J. New developments in molecular orbital theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.

(426) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. A comparison of density functional methods for the estimation of proton chemical shifts with chemical accuracy. *J. Phys. Chem. A* **1999**, *103*, 7357−7363.

(427) Barone, V.; Biczysko, M.; Bloino, J. Fully anharmonic IR and Raman spectra of medium-size molecular systems: accuracy and interpretation. *Phys. Chem. Chem. Phys.* **2014**, *16*, 1759−1787.

(428) Čížek, J. On the Use of the Cluster Expansion and the Technique of Diagrams in Calculations of Correlation Effects in Atoms and Molecules. In *Advances in Chemical Physics*; LeFebvre, R.; Moser, C., Eds.; John Wiley & Sons, Ltd., 1969; Vol. *14*, pp 35−89.

(429) Purvis, G. D.; Bartlett, R. J. A full coupled-cluster singles and doubles model: The inclusion of disconnected triples. *J. Chem. Phys.* **1982**, *76*, 1910−1918.

(430) Scuseria, G. E.; Schaefer, H. F. Is coupled cluster singles and doubles (CCSD) more computationally intensive than quadratic configuration interaction (QCISD)? *J. Chem. Phys.* **1989**, *90*, 3700−3703.

(431) Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F. An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations. *J. Chem. Phys.* **1988**, *89*, 7382−7387.

(432) Lodewyk, M. W.; Soldi, C.; Jones, P. B.; Olmstead, M. M.; Rita, J.; Shaw, J. T.; Tantillo, D. J. The correct structure of aquatolide-experimental validation of a theoretically-predicted structural revision. *J. Am. Chem. Soc.* **2012**, *134*, 18550−18553.

(433) Kutateladze, A. G.; Holt, T. Structure validation of complex natural products: Time to change the paradigm. What did synthesis of Alstofolinine A prove? *J. Org. Chem.* **2019**, *84*, 8297−8299.

(434) Pescitelli, G.; Di Bari, L.; Berova, N. Conformational aspects in the studies of organic compounds by electronic circular dichroism. *Chem. Soc. Rev.* **2011**, *40*, 4603−4625.

(435) Da Silva, H. C.; De Almeida, W. B. Theoretical calculations of 1H NMR chemical shifts for Nitrogenated compounds in Chloroform solution. *Chem. Phys.* **2020**, *528*, 110479.

(436) Saunders, C. M.; Tantillo, D. J. Application of computational chemical shift prediction techniques to the Cereoanhydride structure problem-Carboxylate complications. *Mar. Drugs* **2017**, *15*, 171.

(437) Fornaro, T.; Biczysko, M.; Monti, S.; Barone, V. Dispersion corrected DFT approaches for anharmonic vibrational frequency calculations: Nucleobases and their dimers. *Phys. Chem. Chem. Phys.* **2014**, *16*, 10112−10128.

(438) Zanardi, M.; Marcarino, M. O.; Sarotti, A. M. Redefining the impact of Boltzmann analysis in the stereochemical assignment of polar and flexible molecules by NMR calculations. *Org. Lett.* **2020**, *22*, 52−56.

(439) Kutateladze, A. G.; Reddy, D. S. High-throughput in silico structure validation and revision of halogenated natural products is enabled by parametric corrections to DFT-computed 13C NMR chemical shifts and spin-spin coupling constants. *J. Org. Chem.* **2017**, *82*, 3368−3381.

(440) Sarotti, A. M.; Pellegrinet, S. C. A multi-standard approach for GIAO 13C NMR calculations. *J. Org. Chem.* **2009**, *74*, 7254−7260.

(441) Sarotti, A. M.; Pellegrinet, S. C. Application of the multi-standard methodology for calculating 1H NMR chemical shifts. *J. Org. Chem.* **2012**, *77*, 6059−6065.

(442) Son, W. J.; Jang, S.; Shin, S. Simulated Q-annealing: conformational search with an effective potential. *J. Mol. Model.* **2012**, *18*, 213−220.

(443) Andricioaei, I. I.; Straub, J. E. Generalized simulated annealing algorithms using Tsallis statistics: Application to conformational optimization of a tetrapeptide. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1996**, *53*, R3055−R3058.

(444) Feldman, H. J.; Hogue, C. W. V. Probabilistic sampling of protein conformations: New hope for brute force? *Proteins: Struct., Funct., Genet.* **2002**, *46*, 8−23.

(445) Pracht, P.; Bohle, F.; Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169−7192.

(446) Polavarapu, P. L. Review article molecular structure determination using chiroptical spectroscopy: Where we may go wrong ? *Chirality* **2012**, *24*, 909−920.

(447) *xtb doc: Introduction to CREST*; Grimme Group, 2019.

(448) Navarro-Vázquez, A. When not to rely on Boltzmann populations. Automated CASE-3D structure elucidation of hyacinthacines through chemical shift differences. *Magn. Reson. Chem.* **2029**, *58*, 139−144.

(449) Bootsma, A. N.; Wheeler, S. E. Popular integration grids can result in large errors in DFT-computed free energies. *ChemRxiv* **2019**, 1−18.

(450) Lemonick, S. Density Functional Theory Error Discovered; *Chem. Eng. News* **2019**.*97*.

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

(452) Mazzanti, A.; Casarini, D. Recent trends in conformational analysis. *WIREs: Comput. Mol. Sci.* **2012**, *2*, 613−641.

(453) Sarotti, A. M. Structural revision of two unusual rhamnofolane diterpenes, curcusones I and J, by means of DFT calculations of NMR shifts and coupling constants. *Org. Biomol. Chem.* **2018**, *16*, 944−950.

(454) Tsui, K. Y.; Tombari, R. J.; Olson, D. E.; Tantillo, D. J. Reconsidering the structure of Serlyticin-A. *J. Nat. Prod.* **2019**, *82*, 3464−3468.

(455) Grimblat, N.; Sarotti, A. M. Computational chemistry to the rescue: Modern toolboxes for the assignment of complex molecules by GIAO NMR calculations. *Chem. - Eur. J.* **2016**, *22*, 12246−12261.

(456) Lauro, G.; Das, P.; Riccio, R.; Reddy, D. S.; Bifulco, G. DFT/NMR Approach for the configuration assignment of groups of stereoisomers by the combination and comparison of experimental and predicted sets of data. *J. Org. Chem.* **2020**, *85*, 3297−3306.

(457) Baldridge, K. K.; Siegel, J. S. Correlation of empirical δ(TMS) and absolute NMR chemical shifts predicted by ab initio computations. *J. Phys. Chem. A* **1999**, *103*, 4038−4042.

(458) Gao, P.; Wang, X.; Huang, Z.; Yu, H. 11B NMR chemical shift predictions via density functional theory and gauge-including atomic orbital approach: Applications to structural elucidations of boron-containing molecules. *ACS Omega* **2019**, *4*, 12385−12392.

(459) Saunders, C.; Khaled, M. B.; Weaver, J. D., 3rd; Tantillo, D. J. Prediction of (19)F NMR chemical shifts for fluorinated aromatic compounds. *J. Org. Chem.* **2018**, *83*, 3220−3225.

(460) Latypov, S. K.; Polyancev, F. M.; Yakhvarov, D. G.; Sinyashin, O. G. Quantum chemical calculations of 31P NMR chemical shifts: Scopes and limitations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6976−6987.

(461) Bally, T.; Rablen, P. R. Quantum-chemical simulation of 1H NMR spectra. 2. Comparison of DFT-based procedures for computing proton-proton coupling constants in organic molecules. *J. Org. Chem.* **2011**, *76*, 4818−4830.

(462) Lacerda, E. G., Jr; Kamounah, F. S.; Coutinho, K.; Sauer, S. P. A.; Hansen, P. E.; Hammerich, O. Computational prediction of 1H and 13C NMR chemical shifts for protonated Alkylpyrroles: Electron correlation and not solvation is the salvation. *ChemPhysChem* **2019**, *20*, 78−91.

(463) Hansen, P. E.; Lund, T.; Krake, J.; Spanget-Larsen, J.; Hvidt, S. A Reinvestigation of the ionic liquid Diisopropylethylammonium Formate by NMR and DFT methods. *J. Phys. Chem. B* **2016**, *120*, 11279−11286.

(464) Vázquez, J. T. Tetrahedron: Asymmetry features of electronic circular dichroism and tips for its use in determining absolute configuration. *Tetrahedron: Asymmetry* **2017**, *28*, 1199−1211.

(465) McAlpine, J. B.; Chen, S. N.; Kutateladze, A.; MacMillan, J. B.; Appendino, G.; Barison, A.; Beniddir, M. A.; Biavatti, M. W.; Bluml, S.; Boufridi, A. The value of universally available raw NMR data for transparency, reproducibility, and integrity in natural product research. *Nat. Prod. Rep.* **2019**, *36*, 35−107.

(466) Lanyon, B. P.; Whitfield, J. D.; Gillett, G. G.; Goggin, M. E.; Almeida, M. P.; Kassal, I.; Biamonte, J. D.; Mohseni, M.; Powell, B. J.;

Barbieri, M.; Aspuru-Guzik, A.; White, A. G. Towards quantum chemistry on a quantum computer. *Nat. Chem.* **2010**, *2*, 106−111.

(467) Snyder, K. M.; Sikorska, J.; Ye, T.; Fang, L.; Su, W.; Carter, R. G.; McPhail, K. L.; Cheong, P. H. Y. Towards theory driven structure elucidation of complex natural products: Mandelalides and coibamide A. *Org. Biomol. Chem.* **2016**, *14*, 5826−5831.

(468) Brueckner, A. C.; Ogba, O. M.; Snyder, K. M.; Richardson, H. C.; Cheong, P. H.-Y. Conformational Searching for Complex, Flexible Molecules. In *Applied Theoretical Organic Chemistry*; World Scientific, 2018, Chapter 5, pp 147−164.

(469) Yu, H.; Lin, Y. S. Toward structure prediction of cyclic peptides. *Phys. Chem. Chem. Phys.* **2015**, *17*, 4210−4219.

(470) Nguyen, Q. N. N.; Schwochert, J.; Tantillo, D. J.; Lokey, R. S. Using 1H and 13C NMR chemical shifts to determine cyclic peptide conformations: A combined molecular dynamics and quantum mechanics approach. *Phys. Chem. Chem. Phys.* **2018**, *20*, 14003−14012.

(471) Burns, D. C.; Mazzola, E. P.; Reynolds, W. F. The role of computer-assisted structure elucidation (CASE) programs in the structure elucidation of complex natural products. *Nat. Prod. Rep.* **2019**, *36*, 919−933.

(472) Elyashberg, M.; Williams, A. J.; Blinov, K. Structural revisions of natural products by Computer-Assisted Structure Elucidation (CASE) systems. *Nat. Prod. Rep.* **2010**, *27*, 1296−1328.

(473) Williams, A. J.; Elyashberg, M. E.; Blinov, K. A.; Lankin, D. C.; Martin, G. E.; Reynolds, W. F.; Porco, J. A.; Singleton, C. A.; Su, S. Applying computer-assisted structure elucidation algorithms for the purpose of structure validation: Revisiting the NMR assignments of hexacyclinol. *J. Nat. Prod.* **2008**, *71*, 581−588.

(474) Lindel, T.; Junker, J.; Köck, M. 2D-NMR-guided constitutional analysis of organic compounds employing the computer program COCON[#]. *Eur. J. Org. Chem.* **1999**, *1999*, 573−577.

(475) Lindel, T.; Junker, J.; Köck, M. Cocon: From NMR correlation data to molecular constitutions. *J. Mol. Model.* **1997**, *3*, 364−368.

(476) Nuzillard, J.-M. Automatic structure determination of organic molecules: Principle and implementation of the LSD program. *Chin. J. Chem.* **2003**, *21*, 1263−1267.

(477) Dos, F. M., Jr.; Velozo, L. S.; de Carvalho, E. M.; Marques, A. M.; Borges, R. M.; Trindade, A. P.; dos Santos, M. I.; de Albuquerque, A. C.; Costa, F. L.; Kaplan, M. A.; de Amorim, M. B. 3-Ishwarone, a rare ishwarane sesquiterpene from Peperomia scandens Ruiz & Pavon: structural elucidation through a joint experimental and theoretical study. *Molecules* **2013**, *18*, 13520−13529.

(478) Junior, F. M.; Covington, C. L.; de Albuquerque, A. C.; Lobo, J. F.; Borges, R. M.; de Amorim, M. B.; Polavarapu, P. L. Absolute configuration of (−)-Centratherin, a sesquiterpenoid lactone, defined by means of chiroptical spectroscopy. *J. Nat. Prod.* **2015**, *78*, 2617−2623.

(479) Wang, B.; Dossey, A. T.; Walse, S. S.; Edison, A. S.; Merz, K. M. Relative configuration of natural products using NMR chemical shifts. *J. Nat. Prod.* **2009**, *72*, 709−713.

(480) Palazzo, T. A.; Truong, T. T.; Wong, S. M. T.; Mack, E. T.; Lodewyk, M. W.; Harrison, J. G.; Gamage, R. A.; Siegel, J. B.; Kurth, M. J.; Tantillo, D. J. Reassigning the structures of natural products using NMR chemical shifts computed with quantum mechanics: A laboratory exercise. *J. Chem. Educ.* **2015**, *92*, 561−566.

(481) Banert, K.; Tantillo, D. J. A problem in the structure assignment of acremolin C, which is most probably identical with acremolin B. *Nat. Prod. Res.* **2019**, *33*, 3011−3015.

(482) Fu, T.; Oetjen, J.; Chapelle, M.; Verdu, A.; Szesny, M.; Chaumot, A.; Degli-Esposti, D.; Geffard, O.; Clément, Y.; Salvador, A.; Ayciriex, S. In situ isobaric lipid mapping by MALDI-ion mobility separation-mass spectrometry imaging. *J. Mass Spectrom.* **2020**, *55*, No. e4531.

(483) Jackson, S. N.; Barbacci, D.; Egan, T.; Lewis, E. K.; Schultz, J. A.; Woods, A. S. MALDI-ion mobility mass spectrometry of lipids in negative ion mode. *Anal. Methods* **2014**, *6*, 5001−5007.

(484) Nagy, G.; Veličković, D.; Chu, R. K.; Carrell, A. A.; Weston, D. J.; Ibrahim, Y. M.; Anderton, C. R.; Smith, R. D. Towards resolving the spatial metabolome with unambiguous molecular annotations in complex biological systems by coupling mass spectrometry imaging with structures for lossless ion manipulations. *Chem. Commun.* **2019**, *55*, 306−309.

(485) Neumann, E. K.; Migas, L. G.; Allen, J. L.; Caprioli, R. M.; Van de Plas, R.; Spraggins, J. M. Spatial metabolomics of the human kidney using MALDI trapped ion mobility imaging mass spectrometry. *Anal. Chem.* **2020**, *92*, 13084−13091.

(486) Rivera, E. S.; Djambazova, K. V.; Neumann, E. K.; Caprioli, R. M.; Spraggins, J. M. Integrating ion mobility and imaging mass spectrometry for comprehensive analysis of biological tissues: A brief review and perspective. *J. Mass Spectrom.* **2020**, *55*, e4614.

(487) Stopka, S. A.; Agtuca, B. J.; Koppenaal, D. W.; Paša-Tolić, L.; Stacey, G.; Vertes, A.; Anderton, C. R. Laser-ablation electrospray ionization mass spectrometry with ion mobility separation reveals metabolites in the symbiotic interactions of soybean roots and rhizobia. *Plant J.* **2017**, *91*, 340−354.

(488) Kuhn, S.; Colreavy-Donnelly, S.; Santana de Souza, J.; Borges, R. M. An integrated approach for mixture analysis using MS and NMR techniques. *Faraday Discuss.* **2019**, *218*, 339−353.

(489) Clendinen, C. S.; Stupp, G. S.; Ajredini, R.; Lee-McMullen, B.; Beecher, C.; Edison, A. S. An overview of methods using (13)C for improved compound identification in metabolomics and natural products. *Front. Plant Sci.* **2015**, *6*, 611.

(490) Hogben, H. J.; Krzystyniak, M.; Charnock, G. T. P.; Hore, P. J.; Kuprov, I. Spinach - A software library for simulation of spin dynamics in large spin systems. *J. Magn. Reson.* **2011**, *208*, 179−194.

(491) Kuprov, I. Large-scale NMR simulations in liquid state: A tutorial. *Magn. Reson. Chem.* **2018**, *56*, 415−437.

(492) Bingol, K.; Bruschweiler-Li, L.; Yu, C.; Somogyi, A.; Zhang, F.; Bruschweiler, R. Metabolomics beyond spectroscopic databases: a combined MS/NMR strategy for the rapid identification of new metabolites in complex mixtures. *Anal. Chem.* **2015**, *87*, 3864−3870.

(493) Pence, H. E.; Williams, A. ChemSpider: An online chemical information resource. *J. Chem. Educ.* **2010**, *87*, 1123−1124.

(494) Sanchez-Lengeling, B.; Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **2018**, *361*, 360−365.

(495) Schwalbe-Koda, D.; Gómez-Bombarelli, R. Generative models for automatic chemical design; *arXiv* **2019**.1907.01632.

(496) Xue, D.; Gong, Y.; Yang, Z.; Chuai, G.; Qu, S.; Shen, A.; Yu, J.; Liu, Q. Advances and challenges in deep generative models for de novo molecule generation. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2019**, *9*, No. e1395.

(497) Colby, S. M.; Thomas, D. G.; Nuñez, J. R.; Baxter, D. J.; Glaesemann, K. R.; Brown, J. M.; Pirrung, M. A.; Govind, N.; Teeguarden, J. G.; Metz, T. O.; Renslow, R. S. ISiCLE: A quantum chemistry pipeline for establishing in silico collision cross section libraries. *Anal. Chem.* **2019**, *91*, 4346−4356.

(498) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res.* **2014**, *42*, W94−99.

(499) Paglia, G.; Williams, J. P.; Menikarachchi, L.; Thompson, J. W.; Tyldesley-Worster, R.; Halldorsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; Palsson, B. O.; Astarita, G. Ion mobility derived collision cross sections to support metabolomics applications. *Anal. Chem.* **2014**, *86*, 3985−3993.

(500) Wyttenbach, T.; Bushnell, J. E.; Bowers, M. T. Salt bridge structures in the absence of solvent? The case for the oligoglycines. *J. Am. Chem. Soc.* **1998**, *120*, 5098−5103.

(501) Xia, R.; Kais, S. Quantum machine learning for electronic structure calculations. *Nat. Commun.* **2018**, *9*, 4195.

(502) Schütt, K.; Gastegger, M.; Tkatchenko, A.; Müller, K.-R.; Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.

(503) Häse, F.; Kreisbeck, C.; Aspuru-Guzik, A. Machine learning for quantum dynamics: deep learning of excitation energy transfer properties. *Chemical science* **2017**, *8*, 8419−8426.

(504) Wiebe, N.; Kapoor, A.; Svore, K. M. Quantum deep learning; *arXiv* **2014**.1412.3489v2.

(505) Colby, S. M.; Nuñez, J. R.; Hodas, N. O.; Corley, C. D.; Renslow, R. R. Deep learning to generate in silico chemical property libraries and candidate molecules for small molecule identification in complex samples. *Anal. Chem.* **2020**, *92*, 1720−1729.

(506) Blaschke, T.; Olivecrona, M.; Engkvist, O.; Bajorath, J.; Chen, H. Application of generative autoencoder in de novo molecular design. *Mol. Inf.* **2018**, *37*, 1700123.

(507) Dai, H.; Tian, Y.; Dai, B.; Skiena, S.; Song, L. Syntax-directed variational autoencoder for structured data; *arXiv* **2018**.1802.08786.

(508) De Cao, N.; Kipf, T. MolGAN: An implicit generative model for small molecular graphs; *arXiv* **2018**.1805.11973v1.

(509) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **2018**, *4*, 268−276.

(510) Gupta, A.; Müller, A. T.; Huisman, B. J.; Fuchs, J. A.; Schneider, P.; Schneider, G. Generative recurrent networks for de novo drug design. *Mol. Inf.* **2018**, *37*, 1700111.

(511) Jin, W.; Barzilay, R.; Jaakkola, T. Junction tree variational autoencoder for molecular graph generation; *arXiv* **2019**.1802.04364v4.

(512) Kadurin, A.; Aliper, A.; Kazennov, A.; Mamoshina, P.; Vanhaelen, Q.; Khrabrov, K.; Zhavoronkov, A. The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* **2017**, *8*, 10883.

(513) Kang, S.; Cho, K. Conditional molecular design with deep generative models. *J. Chem. Inf. Model.* **2019**, *59*, 43−52.

(514) Kim, K.; Kang, S.; Yoo, J.; Kwon, Y.; Nam, Y.; Lee, D.; Kim, I.; Choi, Y.-S.; Jung, Y.; Kim, S.; et al. Deep-learning-based inverse design model for intelligent discovery of organic molecules. *npj Comput. Mater.* **2018**, *4*, 67.

(515) Lim, J.; Ryu, S.; Kim, J. W.; Kim, W. Y. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *J. Cheminf.* **2018**, *10*, 31.

(516) Merk, D.; Friedrich, L.; Grisoni, F.; Schneider, G. De novo design of bioactive small molecules by artificial intelligence. *Mol. Inf.* **2018**, *37*, 1700153.

(517) Miyao, T.; Funatsu, K.; Bajorath, J. Exploring differential evolution for inverse QSAR analysis. *F1000Research* **2017**, 6.

(518) Wong, W. W.; Burkowski, F. J. A constructive approach for discovering new drug leads: Using a kernel methodology for the inverse-QSAR problem. *J. Cheminf.* **2009**, *1*, 4.

(519) Schneider, P.; Schneider, G. De novo design at the edge of chaos: Miniperspective. *J. Med. Chem.* **2016**, *59*, 4077−4086.

(520) Dobson, C. M. Chemical space and biology. *Nature* **2004**, *432*, 824−828.

(521) Reymond, J.-L. The chemical space project. *Acc. Chem. Res.* **2015**, *48*, 722−730.

(522) Reymond, J.-L.; Awale, M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.* **2012**, *3*, 649−657.

(523) Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **2012**, *52*, 2864−2875.

(524) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem substance and compound databases. *Nucleic Acids Res.* **2016**, *44*, D1202−D1213.

(525) Gu, J.; Gui, Y.; Chen, L.; Yuan, G.; Lu, H. Z.; Xu, X. Use of natural products as chemical library for drug discovery and network pharmacology. *PLoS One* **2013**, *8*, No. e62839.

(526) Wishart, D. S.; Feunang, Y. D.; Marcu, A.; Guo, A. C.; Liang, K.; Vazquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; Sayeeda, Z.; Lo, E.; Assempour, N.; Berjanskii, M.; Singhal, S.; Arndt, D.; Liang, Y.; Badran, H.; Grant, J.; Serra-Cayuela, A.; et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608−D617.

(527) Richard, A. M.; Williams, C. R. Distributed structure-searchable toxicity (DSSTox) public database network: a proposal. *Mutat. Res., Fundam. Mol. Mech. Mutagen.* **2002**, *499*, 27−52.

(528) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(529) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(530) Stephens, P. J.; Devlin, F.; Chabalowski, C.; Frisch, M. J. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(531) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.

(532) Feller, D. The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* **1996**, *17*, 1571−1586.

(533) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, J.; Windus, T. L. Basis set exchange: a community database for computational sciences. *J. Chem. Inf. Model.* **2007**, *47*, 1045−1052.

(534) Liu, S.; Li, J.; Bennett, K. C.; Ganoe, B.; Stauch, T.; Head-Gordon, M.; Hexemer, A.; Ushizima, D.; Head-Gordon, T. Multi-resolution 3D-DenseNet for Chemical Shift Prediction in NMR Crystallography. *J. Phys. Chem. Lett.* **2019**, *10*, 4558−4565.

(535) Gao, P.; Zhang, J.; Peng, Q.; Zhang, J.; Glezakou, V. A. General protocol for the accurate prediction of molecular (13)C/(1) H NMR chemical shifts via machine learning augmented DFT. *J. Chem. Inf. Model.* **2020**, *60*, 3746.

(536) Brouard, C.; Basse, A.; d'Alche-Buc, F.; Rousu, J. Improved small molecule identification through learning combinations of Kernel regression models. *Metabolites* **2019**, *9*, 160.

(537) Zhou, X. X.; Zeng, W. F.; Chi, H.; Luo, C.; Liu, C.; Zhan, J.; He, S. M.; Zhang, Z. pDeep: Predicting MS/MS spectra of peptides with deep learning. *Anal. Chem.* **2017**, *89*, 12690−12697.

(538) Li, S.; Arnold, R. J.; Tang, H.; Radivojac, P. On the accuracy and limits of peptide fragmentation spectrum prediction. *Anal. Chem.* **2011**, *83*, 790−796.

(539) Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K. K.; Deming, L.; Berndl, M.; Brant, A.; Cimermancic, P.; Cox, J. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nat. Methods* **2019**, *16*, 519−525.

(540) Degroeve, S.; Martens, L. MS2PIP: a tool for MS/MS peak intensity prediction. *Bioinformatics* **2013**, *29*, 3199−3203.

(541) Gessulat, S.; Schmidt, T.; Zolg, D. P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; Reimer, U.; Ehrlich, H. C.; Aiche, S.; Kuster, B.; Wilhelm, M. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **2019**, *16*, 509−518.

(542) Kingma, D. P.; Welling, M. Auto-encoding variational bayes; *arXiv* **2013**.1312.6114v10.