

Collision Cross Section Prediction Based on Machine Learning

Xiaohang Li ^{1,2,†}, Hongda Wang ^{1,2,†} , Meiting Jiang ^{1,2,†}, Mengxiang Ding ^{1,2}, Xiaoyan Xu ^{1,2}, Bei Xu ^{1,2}, Yadan Zou ^{1,2}, Yuetong Yu ¹ and Wenzhi Yang ^{1,2,*}

¹ State Key Laboratory of Component-Based Chinese Medicine, Tianjin University of Traditional Chinese Medicine, 10 Poyanghu Road, Tianjin 301617, China; xiaohang0519@163.com (X.L.); 17862987156@163.com (H.W.); jiangmeiting21@163.com (M.J.); mengxiang98@163.com (M.D.); xxy_0421@163.com (X.X.); fullmoonvn@163.com (B.X.); zouyadan01@163.com (Y.Z.); 13165576725@163.com (Y.Y.)

² Haihe Laboratory of Modern Chinese Medicine, Tianjin University of Traditional Chinese Medicine, 10 Poyanghu Road, Tianjin 301617, China

* Correspondence: wzyang0504@tjutcm.edu.cn; Tel.: +86-022-5979-1833

† These authors contributed equally to this work.

Abstract: Ion mobility-mass spectrometry (IM-MS) is a powerful separation technique providing an additional dimension of separation to support the enhanced separation and characterization of complex components from the tissue metabolome and medicinal herbs. The integration of machine learning (ML) with IM-MS can overcome the barrier to the lack of reference standards, promoting the creation of a large number of proprietary collision cross section (CCS) databases, which help to achieve the rapid, comprehensive, and accurate characterization of the contained chemical components. In this review, advances in CCS prediction using ML in the past 2 decades are summarized. The advantages of ion mobility-mass spectrometers and the commercially available ion mobility technologies with different principles (e.g., time dispersive, confinement and selective release, and space dispersive) are introduced and compared. The general procedures involved in CCS prediction based on ML (acquisition and optimization of the independent and dependent variables, model construction and evaluation, etc.) are highlighted. In addition, quantum chemistry, molecular dynamics, and CCS theoretical calculations are also described. Finally, the applications of CCS prediction in metabolomics, natural products, foods, and the other research fields are reflected.

Keywords: ion mobility-mass spectrometry; collision cross section; machine learning; prediction; molecular descriptor



Citation: Li, X.; Wang, H.; Jiang, M.; Ding, M.; Xu, X.; Xu, B.; Zou, Y.; Yu, Y.; Yang, W. Collision Cross Section Prediction Based on Machine Learning. *Molecules* **2023**, *28*, 4050. <https://doi.org/10.3390/molecules28104050>

Academic Editor: Francesco Cacciola

Received: 13 April 2023

Revised: 10 May 2023

Accepted: 10 May 2023

Published: 12 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Ion mobility spectroscopy (IMS), analogous to a gas-phase electrophoresis technique, enables the separation of compounds on the basis of the differences in the mobility of ions through buffer gases under the action of an electric field [1,2]. The difference in mobility is caused mainly by distinctions between the charge, shape, and size of the molecules, which leads to the differences in drift time [3–5]. This difference can be described by the collision cross section (CCS) value. In general, ions with the lower mass and/or more-compact structures have shorter drift times and lower CCS values; the larger the space volume and/or the higher the mass number, the greater the CCS value. This structural dependency makes the CCS value an important parameter for compound identification. The origin of IMS can be traced back to the X-ray experiments of Thomson and Rutherford in the late 19th century [6], which even predates the study of mass spectrometry (MS) by about 15 years [1]. However, because of the commercialization of ion mobility instruments, their combination had not been popularized until recently. Mobility separation occurs in milliseconds and is compatible with the modern mass spectrometers operating at microsecond scanning speeds [7]. The coupling of IMS and MS can thus provide four-dimensional structural information for component characterization, including t_R , CCS, MS, and MS/MS, thereby

having great potential in reducing false-positive results and improving identification confidence [8,9]. Unfortunately, the strategy of purchasing and measuring a large number of reference standards to obtain the standard CCS values is cost-prohibitive and difficult to implement in most cases. Currently, numerous CCS values can be obtained through theoretical calculation and machine-learning-based prediction without sufficient standards [10,11]. The former usually uses molecular modeling to provide the approximate structure of the molecule and then calculates CCS by simulating the interaction between the drift gas and the analyzed ions [7,12]. These methods are relatively time-consuming and require more professionalism. The latter utilizes a large data set containing the experimentally measured CCS values and structural parameters of the compounds themselves to train, validate, and test the regression models [13]. This method has the advantages of fast calculation speed and high accuracy. At present, many CCS prediction platforms are available, such as MetCCS [10], LipidCCS [14], DeepCCS [15], AllCCS [16], and CCSbase [11], etc. Previous IMS-related reviews have described either the principles of different platforms or the advantages of a specific platform. In this review, we give a comprehensive summary on both the principles and the advantages of different platforms, which can thus lay a foundation for the workflows of machine learning (ML) for CCS prediction. In addition, we focus on the general steps of constructing a CCS database on the basis of using ML algorithms (Figure 1). Notably, along with the ML model, the quantum mechanical (QM) workflows have been developed as well [17]. In this review, the commonly used methods and techniques in various links are summarized, and some practical tips are proposed.

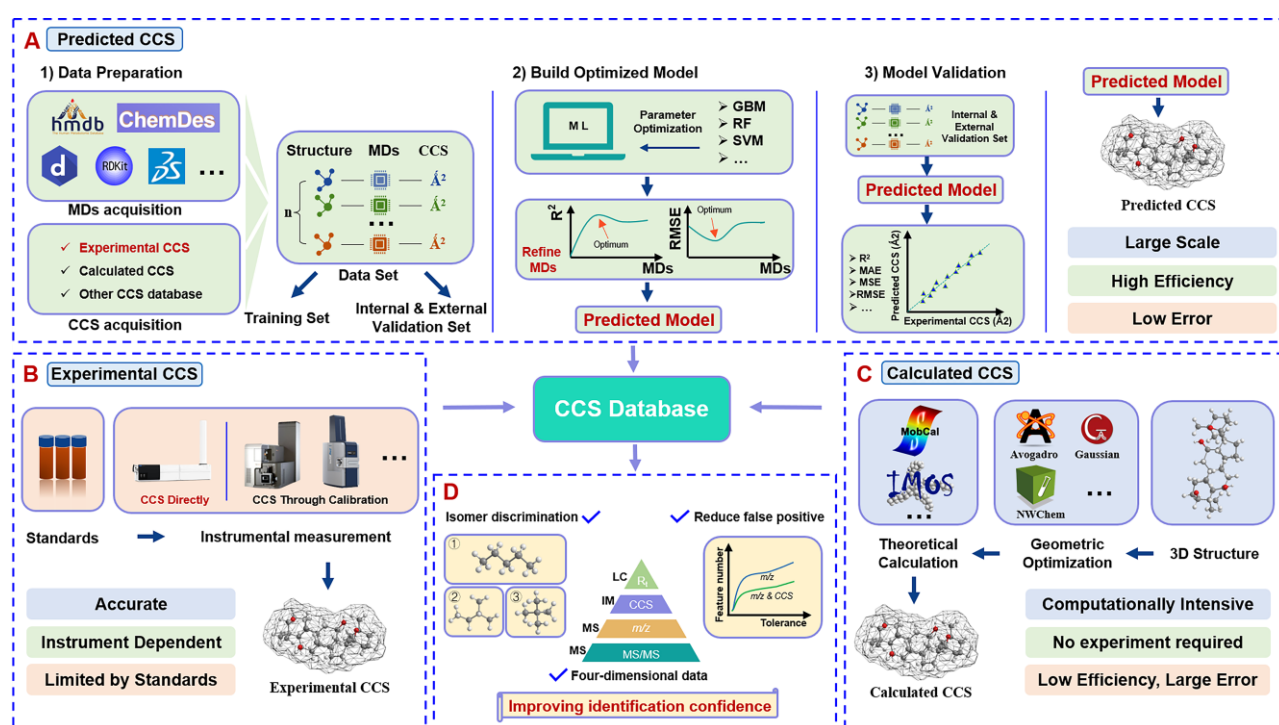


Figure 1. General workflow for building a CCS database: (A) establishing the CCS database on the basis of machine-learning-prediction methods; (B) elaborating the CCS database through ion mobility instrument measurement; (C) creating the CCS database through the theoretical calculation methods; (D) advantages embodied in applying the CCS database for component identification.

2. Ion Mobility-Mass Spectrometry (IM-MS)

2.1. Ion Mobility Platforms with Different Separation Principles

Up to now, the commercially available mainstream IM systems are divided into three types of platforms on the basis of their separation principles: time dispersive, ion confinement with selective release systems, and space dispersive [1,4,18,19]. In current

research, the first two are the most commonly used, and space-dispersive methods have higher development potential. Table 1 shows a comparison of the characteristics of different IM-MS systems.

Table 1. Comparison of commercially available IM-MS techniques.

IMS Technique	Gas State	Resolving Power	Year of Release	CCS Calibration	Available Device	Sort
DTIMS	Stationary	~60–80	2014	Not required	Agilent IM-QTOF	Time dispersive
TWIMS	Stationary	~40–50	2006	Required	Waters Synapt HDMS Waters Vion IMS-QTOF	Time dispersive
SLIMS	Parallel gas flow	~200–300	2021	Required	MOBILion	Time dispersive
TIMS	Parallel gas flow	~200–400	2015	Required	Bruker tims TOF Bruker tims TOF pro Bruker Impact Q-TOF	Confinement and selective release
cIMS	Parallel gas flow	~750	2019	Required	Waters SELECT SERIES cyclic IMS	Confinement and selective release
FAIMS/DMS	Parallel gas flow	Not comparable	2012	-	AB Sciex SelexION	Space dispersive

In a time-dispersive IM-MS system, all ions drift along the same path and are detected by the detector at different times. Generally, ions with small cross-sectional areas are detected first, thanks to their high mobility. Figure 2A shows its specific working principle. The main time-dispersive techniques include drift tube IMS (DTIMS) and traveling-wave IMS (TWIMS). DTIMS consists of several ring electrodes stacked alongside that are filled with an inert static gas through which ions move as directed by a uniform electric field [20]. The drift time can be correlated directly with the CCS value through the Mason–Schamp relationship (Equations (1) and (2)) [21] without requiring a correction program [22,23]. Nevertheless, DTIMS devices have a low-resolution limitation. Researchers have in recent years taken various approaches to improve their resolution, thereby increasing the analysis range and separating isomers that have similar structures, approaches such as increasing the length of the drift tubes [24] to enhance the electric field, introducing multiplexing technology [25–28], developing a new dual drift tube IMS [29], etc. In contrast to DTIMS, ions in TWIMS are directed by a sequence of symmetric potential waves that continuously propagate through the drift region to pass through stationary gases [30,31]. The CCS values of a TWIMS instrument cannot be directly calculated on the basis of the measured drift time, because of the nonuniformly applied electric field. It needs to be calculated on the basis of a group of predefined calibrators, usually using the CCS values derived from DTIMS as a reference [23,32]. Structural similarity between calibrators and analytes is critical for achieving accurate CCS calibration [33,34]. TWIMS has a greater resolution than DTIMS with a uniform electric field for the same drift tube length [5], so TWIMS equipment take up less space while attaining a same resolution level. Recently, structures for lossless ion manipulations (SLIMs), a traveling-wave-based platform, have been developed to guide ions through a printed circuit board path, which maximizes transmission efficiency, increases path length, and achieves an extremely high resolution [35,36]. Cyclic ion mobility-mass spectrometry (cIMS) separates ions in a cyclic mobility chamber and provides significantly longer path lengths by increasing the number of times that ions pass through the cell, thereby improving the resolution and storage capacity of IM separation. The design of the cIMS device allows for IMSⁿ experiments, where ions can undergo multiple instances

of selection, activation, or fragmentation and reseparation before MS detection [37]. The flexibility and practicality of the cIMS separator and control software have led to its wide application in the separation of isomers in different fields [22,38–42]. Time-dispersive instruments allow the simultaneous analysis of all ions and are currently widely used in untargeted metabolomics [31,43,44].

$$\Omega = \frac{3ze}{16NK_0} \sqrt{\frac{2\pi}{\mu k_B T}} \quad (1)$$

$$K_0 = \frac{L}{t_A E} \frac{P}{P_0} \frac{T_0}{T} \quad (2)$$

where Ω is the rotationally averaged CCS, K_0 is the reduced mobility, z is the charge state of the ion, e is the elementary charge, N is the number density of the drift gas, μ is the reduced mass of the ion–neutral drift gas pair, k_B is the Boltzmann constant, T is the gas temperature, t_A is the corrected arrival time, E is the electric field, L is the length of the drift cell, P is the pressure in the drift cell, and P_0 and T_0 are the pressure and the temperature under standard conditions, respectively.

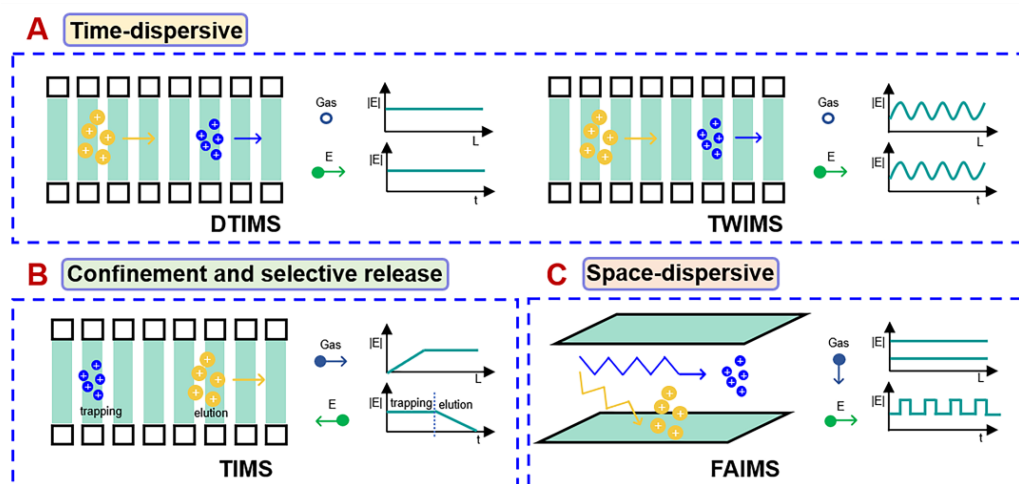


Figure 2. Schematic diagram of drift zone of instruments with different ion mobility values. (A) time dispersive; (B) confinement and selective release; (C) space dispersive. DTIMS: drift tube ion mobility; TWIMS: traveling-wave ion mobility; TIMS: trapped ion mobility; FAIMS: field asymmetric waveform ion mobility.

In a confinement and selective release system, ions are driven by a parallel moving buffer gas and inversely driven by a gradient electric field. When the two forces are equivalent, the ions are stationary relative to the drift tube, indicating that they are trapped. Ions with large cross-sectional areas are stabilized in high-field regions because of their low mobility and the high electric field intensity required to maintain a static state. By reducing the intensity of the electric field, trapped ions are selectively released, and ions with a larger cross-sectional area first pass through the mobility cell and are then detected by the detector [4,45]. Figure 2B shows its specific working principle. Trapped ion mobility spectrometry (TIMS) is the most representative confinement and selective release instrument. It is no longer limited to the length of the device and can provide high resolution three to eight times larger than that of DTIMS or TWIMS [46]. Furthermore, the resolution of TIMS may be modified by adjusting custom parameters, such as the voltage scanning rate (δ) and neutral gas flow rate (vg), making TIMS use very selective [47]. A longer capture time can improve the resolution of the device and ion utilization. Reducing capture time, on the other hand, can enable an untargeted analysis [48]. Importantly, parallel accumulation serial fragmentation (PASEF) can be achieved by connecting two TIMS in series, one for ion

accumulation and the other for ion mobility separation, which improves the duty cycles (up to nearly 100% if equal accumulation and analysis times are used in both TIMS regions) and sensitivity [49], reducing the complexity of the MS/MS spectrum [50]. Like the TWIMS, CCS values cannot be directly determined unless calibration is performed [45,51]. Thanks to its high resolution and sensitivity, TIMS, especially the PASEF strategy based on TIMS, has been applied to the separation of isomers in multiple fields [38,52–54].

The space-dispersive method separates ions along different drift paths on the basis of their mobility in high and low fields, but there is no significant dispersion in time. Figure 2C shows its specific working principle. Field asymmetric waveform ion mobility (FAIMS), also known as differential (ion) mobility spectrometry (DMS or DIMS), belongs to a typical space-dispersive platform [1]. The use of alternating high and low fields in FAIMS forestalls the establishment of a recognized method for obtaining its CCS values [55]. FAIMS acts as a migration filter in which only analytes that have a specific response to changing electric fields and analytes that match the applied compensation voltage can pass through the drift region and the aperture [18,56]. Therefore, FAIMS has been widely used to screen targeted metabolomics and to increase the signal-to-noise ratio of analytes of interest [13].

2.2. Advantages of LC-IM-MS

Recent research has shown that LC-IM-MS has advantages over conventional liquid/gas chromatography-mass spectrometry (LC/GC-MS) in the following four main aspects: (1) providing four-dimensional information to improve the characterization of isomers and enhance the reliability of identification; (2) increasing peak capacity and improving the signal-to-noise ratio (S/N); (3) obtaining additional analysis information when coupling with one or more additional analysis dimensions; (4) improving the quality of spectral acquisition [57,58].

- (1) LC-IM-MS provides four-dimensional information (t_R , CCS, MS, and MS/MS). As a robust parameter for characterization and recognition, CCS provides orthogonal attributes for compound recognition, improving the confidence level of compound annotation [4,59]. IMS technology has proven that it can be used to separate various isomers, such as lipid isomers [60], steroid isomers [61], fatty acid isomers [62], amino acid isomers [22], and carbohydrate isomers [63]. Numerous strategies have been introduced to enhance the IMS characterization of isomers. A combination of chemical derivatization and IMS can improve the detection of steroid isomers [61], metabolites in nicotine [64], and carbohydrates [65]. The integration of dimers or polymers with IM-MS is another effective method for identifying isomers. More accurately predicting the relative differences in CCS between steroid epimers can be achieved through the energy characteristics of the sodium dimer configuration of epimers [66]. The enantiomers of aromatic amino acids can be differentiated by TWIM-MS through their cationization with copper (II) and multimer formation with D-proline (Pro) as a chiral reference compound [67]. The mobility of ions passing through IMS is affected by using different drift gases and/or by doping volatile chiral reagents in drift gases, which can also be used to separate isomers and enantiomers [68,69]. In addition, platforms such as cIMS [42,70–72], multiplexed ion mobility [26,28,73], and TIMS [74] have improved the separation of isomers by improving mobility resolution. IMS can distinguish between conformational isomers [75] and isotopic isomers [22]. By taking into account all relevant errors, *N*-glycan isomers with different conformations can be distinguished on the basis of the CCS gained from the IMS [75]. As we know, lipids have a wide range of structural diversity, with a large number of isomers. A recent study used IMS to analyze the relationship between lipid structure and its gas-phase conformation, providing accurate and comprehensive conformational lipid profiles [76]. IMS has been used in the separation of isomers with different isotopic atomic positions [77] and labeled/unlabeled isotope-substituted isomers [42]. Researchers have found that IMS can be incorporated into the standard LC-MS/MS isotope analysis process as an additional separation mechanism, which can provide

broader separation space and higher identification confidence for metabolic characterization [22].

- (2) Thanks to the advantage of increasing peak capacity and improving the signal-to-noise ratio, IMS can improve the exposure rate of trace components in complex samples [58,78]. Configuring ion mobility technology in MS studies with different ionization principles (ESI, MSI, and MALDI) can increase the peak capacity by at least two times compared with using MS alone [79–81]. It has been reported that when the mass resolution is 35,000 (fwhm), 860 independent ions can be measured, accounting for 15% of the total 5639 counted ions, while the addition of IMS adds 3911 features for signal recognition [79]. Because IMS is used as a separation module between LC and MS, the number of MS features detected in the metabolite composition characterization experiment has significantly increased [82]. IM-MSI can reduce chemical noise and transfer target signals from congested spectral regions, thereby increasing the S/N of metabolites and lipid peaks by nearly 10 times and doubling the image contrast [83]. Some studies have shown that compared to the traditional lipidomics methods, LC-IM-MS analysis has an increased S/N and can detect a low abundance of phospholipids in highly complex brain lipid samples [43]. In the experiment of adding IMS to MS imaging, it was concluded that lipids with different CCS values can be spatially separated, highlighting their spatial positioning and achieving more-accurate lipid recognition [79].
- (3) In addition to IMS's direct use or combining IMS with LC, it can also combine with gas chromatography (GC), mass spectrometry imaging (MSI), or supercritical fluid chromatography (SFC) technologies. As a result, multidimensional analytical information is provided, and the selection of methods increases. IMS and LC can provide orthogonal separation, with IMS separation occurring within milliseconds, and it is compatible with modern MS that is running at microsecond scanning speeds, allowing maximum separation of metabolite ions prior to MS characterization. IMS is often used in series with reverse-phase liquid chromatography (RPLC) [84–86] and hydrophilic interaction liquid chromatography (HILIC) [87–89]. Some researchers have also proposed an offline two-dimensional liquid chromatography coupled with an ion mobility-quadrupole time-of-flight mass spectrometry (2D-LC/IM-QTOF-MS) analysis strategy, achieving a comprehensive characterization of multiple components in traditional Chinese medicine [8,58,90]. In addition, a study that coupled IMS with MSI technology achieved the spatial localization of bile acids in sample tissues [91]. One study integrated ultrahigh performance supercritical fluid chromatography/quadrupole time-of-flight mass spectrometry (UHPSFC/QTOF-MS) and ion mobility spectroscopy/time-of-flight mass spectrometry (IMS/QTOF-MS) to establish a lipid omics platform for CCS measurement, which has improved the analytical performance and recognition reliability of lipids [92].
- (4) IM can improve the overall resolution of the spectrum and obtain high-quality MS¹ and MS² spectra. Double-charged ion clusters make the types of precursors thoroughly complex and can easily generate false positives when annotating MS² data. IM is capable of separating dimers or double-charged ions in a full scan spectrum and generating high-resolution spectra of MS¹ and MS² that are close to the standards [58,84]. Wang [58] used an LC-IM-MS system to comprehensively characterize the multicomponents of compound Danshen dripping pills (CDDPs) and elucidated the advantages of IM. IM can improve the overall resolution of the spectrum of CDDPs and effectively distinguish the doubly charged saponins or the dimers of salvianolic acids, to obtain high-quality MS¹ and MS² spectra and reduce the false positives of multicomponent characterization.

3. Collision Cross Section Value: Dependent Variable of the Model

3.1. Acquisition of CCS Values

Experimental measurements [16,37,93–102] and theoretical calculations are the two main ways to obtain CCS values. The latter can adopt various strategies, including theoretical-driven methods [12,94,103–108] and data-driven methods [7,10,14,15,94,109–111].

The experimental CCS values are obtained by acquiring the mobility data of metabolite standards by using ion mobility platforms (DTIMS, TWIMS, TIMS, etc.) that operate under low field conditions. Because they have different principles, most of them require a dedicated calibration process to determine their CCS values [1]. Currently, the stepped-field method in DTIM, considered as the gold standard for CCS measurement, is the only method that does not require calibration to measure CCS values. Another single-field method requires the use of the linear relational equations constructed by the relationship between the CCS value and the drift time of the calibrants to calculate the CCS values [112]. In TWIMS, it is also necessary to use calibrants with known CCS values to construct a nonlinear calibration curve for both, thereby using this curve and the measured drift time to calculate CCS. The selection of calibrants should meet the following conditions: (1) ensuring good chemical stability; (2) providing wide coverage of m/z and CCS and uniform ion distribution; (3) forming multiple charge states; and (4) being structurally similar to the object to be analyzed [23,32,101,113,114]. At present, polyalanines and Agilent ESI-L low concentration tube mixes are widely used calibrants in DTIMS and TWIMS. Unlike the previous two calibration methods, TIMS uses known mobility (K_0) calibrants to establish a linear relationship between the reciprocal of mobility ($1/K_0$) and voltage, further obtaining the measured ion mobility and finally obtaining a CCS value after conversion. The commonly used calibrants for TIMS include perfluoro-phosphazenes [47], Agilent ESI-L low concentration tube mixes, etc. In addition, the construction of a high-precision CCS database is inseparable from the stable operation of the instrument and the calibration program. During data collection, the performance of the instrument is evaluated and its stability monitored by repeatedly measuring quality-control (QC) samples at intervals of a certain number of injections [10,14]. Some researchers have collected CCS values for a large number of metabolite standards, and CCS databases for one or more types of compounds have been constructed. Table 2 shows these specialized databases, from which we can find the following: (1) three ion mobility platforms, namely DTIMS, TWIMS, and TIMS, are involved, of which DTIMS is the most widely used; (2) various types of compounds, such as metabolites, lipids, biological samples, and drugs or drug analogs, are covered; (3) the number of CCS values obtained from this method is relatively small compared to that from the calculation and speculation method. With the deepening of research, experimental CCS databases continue to increase. However, the number of compounds in the experimental CCS database is always limited because of the limitations of the number of compound standards and ion mobility resolution.

Table 2. List of currently available CCS databases.

Source	Research Object	Number of Compounds	Number of CCS Values	Instrument Platform	Web	Ref.
Experimental CCS	Metabolites	125	209	TWIMS	/	[94]
	Lipids	244	244	TWIMS	/	[79]
	Metabolites and xenobiotics	459	826	DTIMS	http://panomics.pnnl.gov/metabolites/ (accessed on 10 February 2023)	[95]
	Primary metabolites	417	1246	DTIMS	/	[96]
	Steroids	300	1080	TWIMS	/	[97]
	Metabolites	1142	3271	DTIMS	/	[59]

Table 2. Cont.

Source	Research Object	Number of Compounds	Number of CCS Values	Instrument Platform	Web	Ref.
	Metabolites	2193	5119	DTIMS, TWIMS	http://allccs.zhulab.cn/ (accessed on 10 February 2023)	[16]
	Metabolites	510	942	TWIMS	/	[98]
	Bile acids	47	400	DTIMS	/	[99]
	Lipids	/	594	DTIMS	/	[100]
	Lipids	1856	1856	TIMS	/	[48]
	Drug-like compounds and pesticides	~500	~500	DTIMS	/	[101]
	Small molecules	124	124	DTIMS, TWIMS	/	[23]
	Drug or drug-like molecules	1425	1440	TWIMS	/	[102]
	Doping agents	192	192	TWIMS	/	[115]
	Metabolites	112	207	TWIMS	https://massive.ucsd.edu (accessed on 10 February 2023)	[116]
	Metabolites	87	142	TWIMS	/	[117]
	Mycotoxins	53	219	TWIMS	/	[118]
	Lipids	217	456	DTIMS	https://mcleanresearchgroup.shinyapps.io/CCS-Compendium/ (accessed on 10 February 2023)	[76]
	N-glycans	500	500	TWIMS	/	[119]
Calculated CCS	ISiCLE: metabolites	/	~1,000,000	/	metabolomics.pnnll.gov	[12]
	Metabolites	125	205	/	/	[94]
	POMICS	/	/	/	https://www.pomics.org/ (accessed on 10 February 2023)	[120]
Predicted CCS	MetCCS: metabolites	35,203	176,015	DTIMS	http://www.metabolomics-shanghai.org/software.php (accessed on 10 February 2023)	[10]
	LipidCCS: lipids	15,646	63,434	DTIMS	http://www.metabolomics-shanghai.org/LipidCCS/ (accessed on 10 February 2023)	[14]
	AllCCS: metabolites	1,670,596	11,697,711	/	http://allccs.zhulab.cn/ (accessed on 10 February 2023)	[16]
	Pesticide residues	336	336	/	/	[110]
	DeepCCS: metabolites	2400	2400	/	/	[15]
	Sterol lipids	2068	2068	/	/	[111]
	Food contact materials	488	635	TWIMS	/	[109]
	dmCCS: drugs and their metabolites	3286	2068	/	https://CCSbase.net/dmccs_predictions (accessed on 10 February 2023)	[7]
	CCSbase: lipids, metabolites, drugs	4742	7669	DTIMS, TWIMS	https://CCSbase.net (accessed on 10 February 2023)	[11]

Another method of obtaining CCS values is to use computational chemistry tools to obtain theoretical CCS values. The general process of this method is as follows: (1) obtain the three-dimensional (3D) structure and possible conformational forms of compounds; (2) use molecular mechanics, molecular dynamics, quantum chemistry (especially density functional theory, DFT), etc., to screen and optimize the conformations of compounds; and (3) select the appropriate algorithm for calculation. Avogadro [105–107], TINKER [121], Gaussian [105,121], NWChem [107], and SPARTAN'18 [106] are commonly used software programs that can achieve geometric optimization. The calculation of CCS values can be im-

plemented through software such as MobCal [105,121], Collidoscope [122], IMoS [106,107], and IMPACT [106]. MobCal is the most commonly used software, and it provides three algorithms: projection approximation (PA), exact hard sphere scattering (EHSS), and the trajectory method (TM) [103]. PA is the simplest, fastest, and most widely used method that reduces scattering in 3D space to simpler, low-dimensional projections. The molecule is represented as a collection of overlapping hard spheres in PA, and the calculated CCS value is the rotational average of the projected area of this set [121]. The successive introduction of the projected superposition approximation (PSA) method [104] and the local collision probability approximation (LCPA) method [123] solved a problem: the collision between ions and gases as well as noncovalent interactions were not considered in the principle of the PA method. The EHSS method simulates the trajectory of the drift gas approaching and colliding with analyte ions [124]. The algorithm is relatively complex and is often used in the calculation of macromolecular CCS. TM is the most complex and computationally intensive method among the commonly used methods. It simulates the 3D scattering of buffer gas particles under the influence of long-range interaction potential, and it takes into account the van der Waals force and polarization interaction [103]. Based on the TM algorithm, Collidoscope uses parallel computing and trajectory parameter optimization, resulting in a significant reduction in computing time [122]. The underlying algorithm of IMoS [125] is different from that of MobCal and includes the richest CCS computing methods: PA, EHSS, DHSS, TM, and DTM. Table 2 also shows the information from the CCS database obtained through theoretical calculation methods. Zanotto [126] developed high-performance CCS computation software (HPCCS), which performs CCS calculation by using high-performance computing techniques. By using the trajectory method, HPCCS can accurately calculate CCS values for a great variety of molecules, ranging from small organic molecules to large protein complexes, using helium or nitrogen as a buffer gas with considerable gains in computer time compared with publicly available codes under the same level of theory (Table 3). CoSIMS [127] is another CCS computation-software-based multithreaded trajectory method, and it is able to calculate nearly identical CCS values as MobCal can in nearly two orders of magnitude less CPU time thanks to the various numerical methods implemented into the software, even when run on a single CPU core (Table 3). Colby [12] generated a structure and chemical property library by using molecular dynamics, quantum chemistry, and ion mobility calculations and obtained over one million CCS values by using the developed in silico chemical library engine (ISiCLE). This research reconstructed the popular MobCal code for trajectory calculation, improving the computational efficiency by more than two orders of magnitude. The method of obtaining theoretical CCS in computational chemistry has certain limitations, though: (1) a large amount of calculation and logical judgment; (2) low efficiency and a long calculation time (CCS calculation of a compound often takes several days); and (3) a large CCS error, about 3–30% [128]. Therefore, the theoretical calculation accuracy and the efficiency of CCS need to be further improved. Importantly, the accuracy of CCS calculations often depends on a variety of factors, such as different buffer gases in actual measurements and whether they are corrected, the choice of different force fields during conformation generation, and different algorithms for theoretical calculations [129,130].

Table 3. The current CCS computation software.

Software	Year	Methods	Collision Gas	Open Source	Ref.
MobCal	1996	PA, EHSS, TM	He/N ₂	Yes	[131]
IMoS	2013	DTM, DHSS	He/N ₂	Yes	[125]
IMPACT	2015	PA	He	Yes	[132]
Collidoscope	2017	TM	He/N ₂	Yes	[122]
HPCCS	2018	TM	He/N ₂	Yes	[126]
CoSIMS	2019	TM	He	Yes	[127]

CCS values can also be obtained through data-driven ML methods. Developing a CCS prediction model that is based on ML requires three components: a training data set, a prediction algorithm, and a validation data set. The training data set contains parameters representing molecular structural properties and CCS values. There are various ways to reflect molecular structural properties (commonly known as molecular descriptors), and the relevant content will be described in Section 4. The training data set can use experimental or theoretical CCS values, usually using the former. The format of the validation data set should be consistent with the training data set, but the two are independent of each other, and there is no data duplication. ML algorithms are used to construct a regression relationship between the molecular structure and the CCS values and are divided into linear and nonlinear methods, which will be described in Section 5. The general process of data modeling includes (1) the acquisition of data sets (randomly divided into a training data set and an internal validation data set, according to a certain proportion); (2) the construction of prediction models (model training, model accuracy evaluation, and model parameter optimization); (3) the validation of prediction models. Different ML models are commonly used to predict the respective CCS values of small molecule compounds and have been applied to metabolites [10,15,16], lipids [14,60], drugs [7], and food [109] and in some other fields [110,111]. It has the following advantages: (1) large prediction scale; (2) fast computing speed without consuming plenty of computing resources; and (3) a small prediction error, 1–3% [128]. Table 2 shows the information obtained from the CCS database through ML algorithms. In addition, ML prediction can be combined with computational chemistry for CCS calculation. For example, Das et al. [120] developed an efficient computational CCS workflow by using the ML model in conjunction with standard DFT methods and CCS calculations. The CCS computation protocols for the calculation of CCS were the following: the determination of the molecular state; conformation generation; conformation filtering; clustering the conformations; DFT geometry optimization; atomic charge calculations; CCS calculation; Boltzmann weighted CCS; and a predicted structure. The complete workflows could make the computation of CCS values tractable for a large number of conformationally flexible metabolites with complex molecular structures.

3.2. Stability Evaluation of CCS Values

As a physicochemical property of chemical compounds, the CCS value has high reproducibility and stability.

- (1) CCS values are consistent among instruments and laboratories. Numerous studies [79,95,99,117,118] have demonstrated that the measurement of CCS values for metabolites with different molecular weights on multiple TWIMS in independent laboratories (between different Vion IMSs and different SynaptG2 HDMSs, as well as between Vion IMS and SynaptG2 HDMS) is repeatable, with an RSD of CCS values within $\pm 3\%$. Sarah [112] studied the reproducibility of CCS values obtained from DTIMS. Upon the completion of the analysis of 51 biologically related standards (amino acids and lipids), it was found that the interlaboratory RSD was $0.30 \pm 0.16\%$. Some studies [23,133] have compared the CCS values measured by TWIMS and DTIMS and found that the absolute percentage error (APE) of the CCS values was within 2%.
- (2) CCS has stability in different substrates. Giuseppe [94] found through experimental measurements that 97% of CCS values had a mean RSD of less than 2%, which demonstrates the repeatability of CCS values in various biological matrices. To test the accuracy and precision of CCS measurements in different matrices, one study [79] compared the CCS values in the database with CCS values measured from a series of lipid extracts such as porcine brain, *E. coli*, and yeast. The results showed that CCS measurements were highly stable in different matrices.
- (3) CCS values have long-term robustness. One study [117] evaluated the reproducibility of the CCS values of steroid compounds after 1.5 years, and the results showed that 95.7% of the CCS values had an RSD within $\pm 1.0\%$.

- (4) CCS also has stability at different sample concentrations. In addition, some studies have proposed some insights into how to improve the repeatability of CCS measurements, especially the high reproducibility between different ion mobility platforms [1,19,134]. For example, consistent instruments, configurations, calibration procedures, etc. are used to achieve measurement standardization; the physical theory behind ion mobility is improved so that different platforms can provide the same, physically correctly calculated CCS values without requiring calibration.

4. Molecular Descriptors: Independent Variable of the Model

4.1. Molecular Representation

Molecular descriptors (MDs) are mathematical representations of molecules calculated by a specific algorithm that converts molecular structures into numbers. MDs can be divided into (1) measured values, such as polarity, log P , molar refractivity, dipole moment, etc., and (2) theoretical values, which can be subdivided into constitutional, topological, geometric, electronic, and physical chemistry types [135]. In addition, there are classification methods for dividing MDs on the basis of different aspects. For example, MDs can also be divided into zero- to three-dimensional descriptors [136]. In research based on ML to predict CCS databases, MDs are often used for prediction [10,14,92,109], and molecular fingerprints (MFs) [137], and molecular quantum numbers (MQNs) [11] have also been used in some studies. MFs, which are included in MDs and are usually in the form of bit vectors, have the advantages of simple operation, fast calculation speed, and high accuracy [138]. However, because of the difficulty in variable selection, there are currently few studies applied to CCS database prediction. Yang [137] creatively used molecular fingerprints and random forest algorithms to conduct CCS prediction research and obtained a CCS database with accurate prediction capabilities ($R^2 = 0.95$, MRE = 2.2%). The MQN system defines a simple and universal chemical space to classify organic molecules and calculate their basic characteristics, including atomic and bond types, polar groups, and topological characteristics [139]. Another study [11] found that using unsupervised clustering based on MQN to decompose chemical structure diversity can train specific and accurate prediction models for each cluster, which showed better performance than using a single model for all data training. This study has broken the limitations of the “black box” prediction model and provides interpretable results. In addition, the quantum-chemical electron ionization mass spectra (QCEIMS/QCxMS) program is the first standalone MD-based program that can predict mass spectra solely on the basis of using molecular structures as inputs [17].

4.2. Access to Molecular Descriptors

MDs can be obtained through specialized computing software, software that includes computing MD functionality and open-source databases or algorithms. Specialized computing software includes PaDEL-Descriptor [140,141], Dragon [142,143], alvaDesc [136,144,145], Mordred [146,147], BlueDesc [145], Chemopy [148], and ChemDesc [149]. Software Discovery Studio [150] includes the calculation of MD functions. The human metabolome database (HMDB) [10], CDK [151,152], RDkit [153], and “rcdk” package [14,60] are open-source databases or algorithms commonly used. Table 4 shows a detailed comparison of some MD acquisition approaches. Thanks to the ability to provide multiple interfaces, such as a graphical user interface (GUI) and a command line interface (CLI), and the ability to calculate plenty of MDs in parallel, PaDEL-Descriptor has become one of the best choices for open-source MD computing [146]. Dragon is another widely used software program for computing MDs. Dragon can calculate a large number of MDs and allows the calculation of disconnected structures (such as salts, complexes, etc.). Although the source code of Dragon is not open, a free and easy-to-use web version of MD (e-Dragon) computing has been developed on the basis of the older version of the software (Dragon 5.4). Further, alvaDesc software can handle full and partial connection structures, provide different unsupervised variable reduction methods, and delete descriptors with constant or missing values to reduce the number of variables, and it can conveniently divide the 33 types

of provided MDs into 2D and 3D ones [109,144,145]. Mordred software can calculate a large number of MDs, and its calculation speed is twice that of PaDEL-Descriptor [146]. BlueDesc can output results in a libSVM input file format, making it easy to build SVM models. ChemoPy is a free software program to calculate 2D and 3D descriptors and can calculate 1135 descriptors. Currently, some web-based MD computing platforms have been developed, such as ChemDes and the Online Chemical Modeling Environment (OCHEM). ChemDes integrates multiple software packages such as CDK, RDKit, and BlueDesc, and it has the functions of structural optimization, molecular format conversion, and similarity calculation. OCHEM is an online version of alvaDesc [154].

Table 4. Comparison between features of MD calculation software programs.

Software	Operating System	Number of Descriptors	Features	Ref.
PaDEL-Descriptor	Windows, Linux, MacOS	>1700	Supports more than 90 molecular file formats	[140]
alvaDesc	Windows, Linux, MacOS	5666	Can handle full and non-full connection structures	[144]
OCHEM	Web	5666	Is a web version of alvaDesc	[154]
chemDes	Web	3679	Integrates with multiple advanced software packages	[149]
Dragon	Windows, Linux, web (e-Dragon)	5270	Has a fast calculation speed, allowing disconnected structures	[a]
Mordred	Windows, Linux, MacOS	>1800	Can calculate macromolecule descriptor	[146]
BlueDesc	Windows, Linux, MacOS	174	Is only applicable to 3D structures	[b]
Chemopy	Windows, Linux	1135	Is applicable to 2D and 3D structures	[148]
Discovery Studio	Windows, Linux	Hundreds	Enables structural optimization	[150]
CDK	Development kit	268	Contains the chemical and bioinformatics Java library	[151]
RDkit	Development kit	200	Is based on the Python language, supporting multiple file formats	[153]
rcdk	Development kit	221	Has the CDK toolkit integrated under the R language	[106]

[a]: http://www.taletе.mi.it/products/dragon_description.htm (accessed on 10 February 2023). [b]: http://www.ra.cs.uni-tuebingen.de/software/bluedesc/welcome_e.html (accessed on 10 February 2023).

4.3. Preprocessing and Optimization of Molecular Descriptors

The main two points that generally suitable MDs should meet are as follows: (1) the correlation between MDs should be as low as possible, and (2) they should have a good correlation with one or more properties of molecules. To accurately reflect the structure of molecules, 2D absolute configurations or optimized 3D configurations should be obtained before calculating the MDs. The 2D absolute configuration of the obtained compound can be minimized by using the MM2 method in Chem3D Ultra software to minimize the energy of the chemical structure of the molecule [150], and thus a stable molecular conformation can be obtained. After obtaining the MDs, reducing their complexity and optimizing their type and quantity, especially for compounds with similar chemical structures (such as lipids), are prerequisites for obtaining high-precision CCS value predictions. Relevant research [14] has found that through comparison, the prediction accuracy of the CCS values of optimized lipid MDs has been greatly improved ($R^2 = 0.9941$, and $R^2 = 0.1322$ before optimization) and the common overfitting problem in lipid prediction has also been solved. The general process of MD optimization is as follows: (1) remove the same values [60],

zero values, and missing values in the data set [144]; (2) eliminate a portion of the MDs that are highly correlated [148]; and (3) gradually remove the MDs that contribute little to the regression model [14,145]. Specifically, the related MDs in the third point can be deleted by using the `nearZeroVar` function in the R package `insert` [109]. Some studies [110] have used the sensitivity analysis techniques in Alyuda NeuroIntelligence software to analyze the importance of the obtained MDs, and they ultimately obtained good CCS value prediction results (with a median relative error of less than 2%). The importance of MDs is calculated by the degree of degradation of model performance after removing the MDs. In one study [145], in extreme gradient-boosting models, the contribution of each variable to the model is calculated on the basis of the number of times that it is selected for splitting, and the square of the improvement to the model is weighted by each split. The deletion or retention of MDs is determined on the basis of their importance to the model. In order to obtain high-precision prediction results, researchers have made efforts to use a combination of 2D descriptors and new 3D descriptors [7], optimizing 3D descriptors [155], and considering the ionization states of protonated and deprotonated sites [12,145,156]. The overall trend is that the compounds used to calculate MDs are closer to the true ionization state. However, some studies [143] have found that the prediction results of 3D models are superior to 2D models in only a few cases, by comparing the impact of using 2D and 3D MDs on predicting CCS performance. Therefore, it is believed that 3D energy minimization structures are usually time-consuming, hindering the realization of high throughput [142].

5. Machine-Learning Algorithms

5.1. Different Prediction Algorithms and Prediction Platforms

Prediction algorithms are used to establish a correlation between the structure of molecules and CCS values and are divided mainly into linear and nonlinear methods (Table 5). Linear modeling methods include stepwise multiple linear analysis (SMLR), principal component regression (PCR), partial least squares (PLS) regression, and the least absolute shrinkage and selection operator (LASSO) algorithm. Common nonlinear algorithms include support vector machine (SVM), neural networks, random forest (RF), and a gradient-boosting machine (GBM).

One study [156] explored the use of MDs and chemometrics tools, namely SMLR, PCR, and PLS regression, to establish predictive models for the respective CCS values of deprotonated phenolic compounds. These methods can be used in routine metabolite identification analysis. Soper-Hopper [142] used the PLS toolbox in Matlab to conduct a PLS analysis of MD and CCS values. The results showed that by using the PLS regression model of MDs, accurate CCS values can be predicted from 2D structural information. Wang [60] developed a method based on the LASSO algorithm to predict the CCS value of lipids. In this method, a series of MDs were screened and optimized to reflect the subtle structural differences between different lipid isomers. The use of MDs and a large number of standard CCS values for lipids has significantly improved the accuracy of the LASSO model. The accuracy was externally verified by using an independent data set, with median relative errors (MREs) of <1.1%. Compared with linear regression algorithms, nonlinear modeling methods have been studied more widely. The following sections will mainly introduce the commonly used nonlinear algorithms for CCS database prediction.

SVR uses a kernel function to map the MDs of metabolites into a high-dimensional feature space, establish a hyperplane in this space, and perform high-dimensional regression between the MDs and CCS values in the training data set [10]. In order to obtain high-precision CSS value prediction results, training data sets are used to optimize the kernel function parameters of the regression hyperplane. The cost of constraints navigation (C) and gamma (γ) are important parameters for constant optimization. The mean absolute error (MAE), median absolute error (MDAE), median relative error (MDRE), and root mean square error (RMSE) are used as the calculation performance indicators [10,14,157]. SVR-based prediction can be achieved through the R package “e1071” (<https://cran.r-project.org/web/packages/e1071>, accessed on 20 February 2023) or CCSP

2.0 platform [147,158]. Zhou [10] reported for the first time a MetCCS database using the support vector regression (SVR) algorithm. This study conducted large-scale CCS predictions for 35,203 metabolites in the HMDB. Next, for the study of lipids, a stepwise elimination method was used to screen out 45 MDs that were highly correlated with CCS values. The SVR method was also used to develop a prediction CCS database containing 15,646 lipids, namely LipidCCS, with significantly improved prediction accuracy (MRE = 1%) [14]. Finally, [16] they collected more than 5000 experimental CCS values from 14 experimental data sets as a large-scale training set, and they continued to use the SVR algorithm to develop the world's largest CCS database of different types of small molecule compounds (more than 1.6 million small molecules), which was named AllCCS.

A neural network, also known as an artificial neural network (ANN), is a type of ML. Its name and its structure are inspired by the human brain and simulate the way that biological neuron signals communicate with each other. The neural network consists of a node layer, including an input layer; one or more hidden layers; and an output layer. Each node is connected to another node and has associated weights and thresholds. If the output of any one node is higher than the specified threshold, the node will be activated and send data to the next layer of the network. The deep neural network (DNN) can be understood as a neural network with many hidden layers. A convolutional neural network (CNN) is a subtype of DNNs, consisting of a feature learning section and a prediction section [15]. It learns the internal representation of input through a series of convolution and maximum pooling steps. This internal representation is then used as an input to the multilayer perceptron to perform the prediction. CCS value prediction based on neural network algorithms can be performed on Alyuda NeuroIntelligence 2.2 software [110,133] or built using the Keras library and Tensorflow backend on the programming software Python [15,159]. Pier-Luc [15] established a neural network between the SMILES format and the CCS of compounds, successfully developed a CCS database called DeepCCS on the basis of CNNs, and predicted the CCS values of more than 2400 compounds (MDRE = 2.7%). Colby's research team developed an algorithm, DarkChem, for metabolomics and predicting the CCS values of untargeted small molecules that is based on neural networks [159]. The algorithm used the SMILES format representing the structure of compounds as inputs and extracted CCS values and *m/z* data on compounds from the PubChem database and the ISiCLE database obtained through computational methods as output. A neural network was established to predict the various physical and chemical properties of compounds, including the CCS values. Through this training mechanism, DarkChem can predict CCS with an average error of 2.5% and can predict CCS values of nearly 600,000 small molecules.

GBM is an integrated learning method. "Boosting" refers to an iterative process that integrates multiple individual learners to form a series of weak learners into strong learners, thereby reducing model generalization errors and improving model prediction accuracy. It can be used for mathematical problems such as classification and regression [160]. At the same time, gradient boosting is mostly constructed by decision trees, also known as gradient-boosting decision trees, which have good fitting ability for linear and nonlinear data, can handle continuous and discrete data, and have high prediction accuracy and strong generalization ability. Extreme gradient boosting (XGBoost) is a scalable ML system for tree boosting, featuring efficiency and flexibility [161]. There are a few studies in which GBM algorithms is used to obtain predictive CCS databases. Nye [98] used the GBM algorithm to predict the CCS values of metabolites in their study of comparing the CCS values obtained through TWIMS and UHPLC-IMS. Connelly et al. [162] compared the experimental, theoretical, and predicted CCS values through ML for isomeric drug metabolites. The CCS value predicted by ML was obtained by using the gradient elevator algorithm, and the final prediction accuracy reached up to 2.4%. In a study by Corey [153], nearly 7325 experimental CCS values from 3775 compounds were used as dependent variables, and a prediction model for CCS values was established by using the GBM algorithm. To prevent overfitting, a nested cross-validation strategy was also used in the study. The final model value showed a mean absolute deviation of 1.2% for the data

set outside the sample. Song et al. [145] compared the impact of XGBoost and the SVM algorithm on prediction models in their research on building a database of chemicals related to plastic packaging. It was found that SVM models based on CDK descriptors provided more-accurate prediction results.

Random forest (RF) is a classifier that uses multiple decision tree units to train and predict samples. It was first proposed and developed by Leo Breiman and Adele Cutler [163] and is also an integrated learning algorithm. Unlike GBM, RF uses the bagging idea, which means that the training sets of decision trees are independent of each other, and the decision trees that makeup RF can be generated in parallel with each other, which applies to both the classification and the regression problems. The RF algorithm can be implemented on the R language open-source software package randomForest (v4.6-14) [164]. The research by Ieritano found that the RF regression algorithm showed the best performance in the correlation between differential mobility and CCS values, compared to the DNN model [165]. The average absolute percentage error of the predicted CCS by RF was $2.6 \pm 0.4\%$ for analytes outside the training set. Fan Yang [137] creatively developed a cross-platform CCS value prediction method using RF algorithms and molecular fingerprints. The test accuracy of this model is above 0.85, and the median of the relative residual is around 2.2%.

Table 5. The algorithms for CCS prediction.

Algorithm	Method Type	Tools	Features	Refs.
Stepwise multiple linear analysis (SMLR)	linear	R package <i>MLRMPA</i>	Data need to be normalized to reduce the impact of overfitting	[156]
Principal component regression (PCR)	linear	R package <i>MASS</i>	Can reduce the dimensionality of the data set while maintaining the features with the maximum variance contribution in the data set	[156]
Partial least squares regression (PLS)	linear	Matlab with the PLS toolbox/R package <i>pls</i>	Not sensitive to multicollinearity issues caused by the use of simple linear regression models	[150,156]
Least absolute shrinkage and selection operator (LASSO)	linear	Open-source R programming	Have powerful ability to perform both variable selection and regularization	[60]
Support vector machine (SVM)	nonlinear	R package <i>e1071</i>	Wide application; relatively small sample size; can effectively avoid overfitting	[10,14,16,145,147,158]
Artificial neural network (ANN)	nonlinear	Alyuda NeuroIntelligence 2.2	Can perform supervised learning, unsupervised learning, and semisupervised learning	[15,110,133,159]
Random forest (RF)	nonlinear	The scikit-learn Python package	Low variance; low susceptibility to overfitting; poor model applicability	[137,165]
Gradient-boosting machine (GBM)	nonlinear	XGBoost library	Overfitting often occurs	[98,152,153,162]

5.2. Evaluation and Verification of Prediction Algorithms

The evaluation and the validation of prediction algorithms often use internal and external validation. The data set for internal validation and the training set are from the same instrument, while the external validation set uses different instruments to obtain the CCS values in the data set [7,10]. When internal and external validations are performed on the created CCS prediction model, the decision coefficient R^2 , mean absolute error (MAE), median absolute error (MDAE), mean squared error (MSE), and root mean square error

(RMSE) are used mainly as evaluation indicators for the prediction performance of different models. Their calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\text{CCS} - \widehat{\text{CCS}})^2}{\sum_{i=1}^n (\text{CCS} - \overline{\text{CCS}})^2} \quad (3)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |\text{CCS} - \widehat{\text{CCS}}|}{n} \quad (4)$$

$$\text{MedAE} = \text{median}(|\text{CCS}_1 - \widehat{\text{CCS}}_1|, \dots, |\text{CCS}_n - \widehat{\text{CCS}}_n|) \quad (5)$$

$$\text{MSE} = \frac{\sum_{i=1}^n (\text{CCS} - \widehat{\text{CCS}})^2}{n} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{CCS} - \widehat{\text{CCS}})^2}{n}} \quad (7)$$

where CCS represents the measured CCS value of the compound measured in LC-MS, $\widehat{\text{CCS}}$ represents the predicted CCS value (by ML) of the compound predicted by the constructed model, $\overline{\text{CCS}}$ represents the CCS mean value of a training or verification set, and n refers to the number of samples in the training or verification set. The value of R^2 is between 0 and 1, and the larger the value, the better. The larger the R^2 of the training set, the higher the degree of fitting of the model, and the larger the R^2 of the verification set, the better the prediction ability; the smaller the values of MAE, MDAE, MSE, and RMSE, the more accurate the prediction results of the model and the smaller the error.

6. CCS Prediction Applications

Thanks to the advantages of CCS prediction, some ML-based CCS databases have emerged one after another. These databases and self-built databases have been applied to fields such as metabolomics, natural products, food, and the other research fields (Section 6.4). Figure 3 shows the specific applications and advantages of CCS prediction.

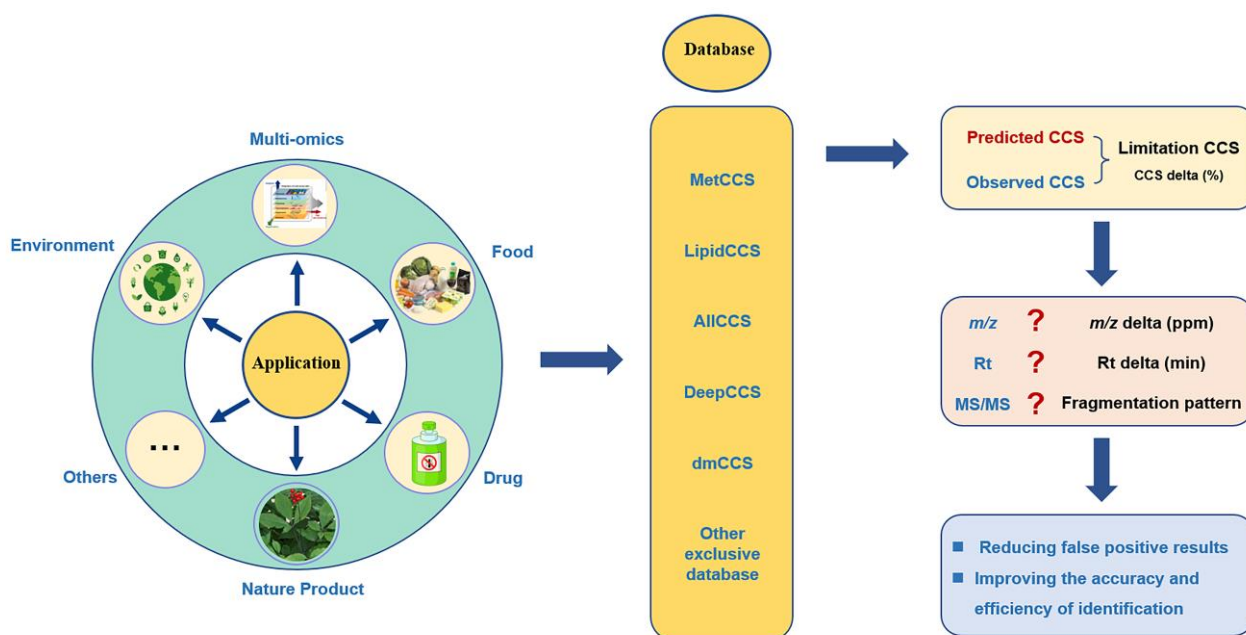


Figure 3. Applications and advantages of CCS prediction.

6.1. In Multiomics

CCS prediction methods based on ML have been widely used in lipidomics, proteomics, and metabolomics. In 2016, Zhou et al. [10] first proposed a strategy for the large-scale calculation of metabolite CCS values using ML methods. They focused on small molecules and used the SVR algorithm to construct a regression relationship between 14 MDs and 400 measured metabolite CCS values. This study ultimately established a predictive CCS database containing 35,203 metabolites, which has high predictive accuracy. The database has also been proven to effectively improve the accuracy and efficiency of identification in untargeted metabolomics. Zhou et al. [14] then used a similar method to construct a regression relationship between the optimized molecular descriptor and more than 450 measured lipid CCS values, and they obtained a CCS value database containing more than 60,000 lipids. Notably, thanks to the high similarity among lipid structures, they used the bioinformatics methods to optimize a set of molecular descriptors and finally established a lipid CCS prediction model with high prediction accuracy. They also concluded that using the database can effectively reduce false-positive lipid identification results in untargeted lipidomics. To annotate both known and unknown metabolites in untargeted metabolomics on the basis of using IM-MS, Zhou et al. [16] developed an integrated multidimensional matching strategy. This strategy integrates over 5000 experimental CCS values and approximately 12 million CCS values predicted by ML, forming a diverse CCS database called AllCCS. The prediction method includes an optimized ML prediction algorithm, a large training data set with a high structure diversity, and a predictive performance evaluation system with representative structure similarity (RSS) score. The AllCCS database has proven to help expand the chemical coverage of identification and reveal comprehensive chemical and metabolic insights into biological processes. The DeepCCS database built by Plante and using convolutional neural network algorithms is trained and validated by using the experimental data sets of over 2400 molecules [15]. Users only need to input the SMILES symbol and the ion type of the compound to easily and quickly obtain the CCS value, which avoids the error problems that users often encounter when using MDs. DeepCCS has been proven to have high prediction accuracy, with a coefficient of determination of 0.97 and a median relative error of 2.7% over a wide molecular range. Wang [60] applied ML prediction to untargeted lipidomics, successfully predicted the CCS values of lipids, and distinguished lipid isomers, including cis–trans isomers. Specifically, a prediction method based on LASSO has been developed and used, and the molecular descriptors of lipids have also been optimized to reflect the subtle differences between their structures. Recently, Rainey [53] reported a high-precision ML algorithm (CCSP 2.0) developed on the basis of SVR models. In particular, CCSP used the open-source Mordred package to calculate a more comprehensive set of MDs. This algorithm can effectively filter false-positive results in metabolomics. Liu [9] developed a quantitative structure–retention relationship (QSRR) strategy and established a 4D information database containing t_R , CCS, MS, and MS/MS for 170 important signaling lipids (*N*-acetyl ethanolamines, NAEs) by using the AllCCS database. Combining it with this database, they identified 68 NAE lipids in different biological samples.

6.2. In Natural Products

The method of CCS prediction can be used to characterize different chemical components in natural products. Song et al. [166] used LC-IM-HDMS techniques to characterize phenolic compounds in bearberry leaves. In this study, a strategy of comparing CCS values obtained from the literature and a database based on ML algorithms (AllCCS) with measured CCS values was added to the component identification workflow, and 88 compounds with high confidence were identified. In their study, a tolerance of 5% between CCS values measured by ML and those predicted by ML was considered acceptable. Wang [167] applied the prediction database (AllCCS) to the component characterization of Chinese traditional medicine *Cuscuta chinensis*. The CCS value predicted by ML provides more possibilities for distinguishing isomers in the absence of reference standards, with a total

of 302 compounds identified or initially identified, of which 109 were not reported. With the continuous expansion of the prediction range and improvement of accuracy in the CCS database, its applications in the component characterization of natural products are becoming increasingly widespread [168,169].

6.3. In Foods

Using the SVM model, Song et al. [109] constructed a correlation between the MDs of 400 food contact materials and the experimentally measured CCS values. In this study, MDs and ML algorithms were optimized, and more-accurate prediction results were obtained. In the meanwhile, they evaluated the applicability of CCS values predicted by ML in the field of food packaging materials by comparing the established CCS value database for food packaging materials with three available predictive CCS values (CCSondemand, AllCCS, and CCSbase), and they found that the prediction given by CCSondemand was the most accurate. This model was eventually applied to the structural annotation of oligomers in polyamide adhesives. By combining it with the self-built prediction CCS database, the recognition confidence of 11 oligomers has been improved. Another study [118] compared the measured CCS values of mycotoxins with two ML databases (AllCCS and CCSbase). The results showed that the CCS values predicted by ML were highly correlated with the measured CCS values (Pearson $r > 0.98$). In the AllCCS prediction model, the prediction error for 91% of the compounds was within a percentage difference of $\pm 5\%$. The CCSbase prediction model provided more-comprehensive structural coverage, resulting in lower deviations, where half of the analytes (50.3%) showed prediction errors within $\pm 2\%$. The above research shows that the use of predicted CCS databases has a certain degree of credibility, which is helpful for the detection of hazardous compounds in foods. Through publicly available CCS databases, it is possible to gain a deeper understanding of the chemical components in food and its contact materials, thereby improving the effectiveness of food safety control.

6.4. In Other Fields

In the field of drugs and drug metabolites, research has used 2D and 3D combined MDs and established large-scale databases to train CCS prediction models on the basis of using ML, achieving high prediction accuracy [7]. Further, 3D information can predict different polymers, conformational isomers, and positional isomers. In a study on the characterization of pesticide components [110], researchers developed an accurate small molecule CCS value prediction tool that was based on ANN and empirical CCS values of 205 organic compounds. The applications of this prediction model to spinach samples have demonstrated its potential for application, which raises confidence in the preliminary identification of suspicious and untargeted pesticides. In environmental testing, Song et al.'s research [145] collected over 1000 experimental CCS values related to plastics in the literature and developed a plastic-packaging database based on SVM models. They applied this CCS database to the identification of plastic-related chemicals in rivers, reducing false-positive results and improving the recognition confidence level. In a compound identification of dust samples [170], the researchers referred to two CCS databases constructed on the basis of using ML during the identification process and evaluated the potential of predictive databases to increase the reliability of compound identification. The applications of predicted CCS values are summarized in Table 6.

Table 6. The applications of predicted CCS values.

Object	Year	Effect	Ref.
Metabolites	2016	MRE < 3%; the identification accuracy can be improved	[10]
	2017	MRE < 1%; the false-positive identifications of lipids can be effectively reduced	[14]
	2019	MRE < 3%; only SMILES notation and ion type are needed	[15]
	2020	MRE < 2%; the accuracy and coverage of both known metabolite and unknown metabolite annotation from biological samples can be improved	[16]
	2022	MRE < 1.1%; <i>cis-trans</i> and sn-positional isomers can be distinguished	[60]
	2022	MRE < 2%; the false positives can be filtered out	[147]
Natural products	2021	a higher identification confidence level can be obtained	[166]
	2022	more possibilities to distinguish isomers can be provided	[167]
Foods	2020	a certain degree of credibility can be obtained	[118]
	2022	MRE < 2%; the identification confidence of 11 oligomers can be improved	[109]
Drugs	2017	MRE < 2%; the confidence in the tentative identification of suspect and nontarget pesticides can be notably improved	[110]
	2022	MRE < 2.2%; sufficient precision to differentiate isomers and conformers can be obtained	[7]
Environment	2020	identification confidence can be increased	[170]
	2022	MRE < 2%; the false positives were reduced, and the recognition confidence levels can be improved	[145]

7. Summary and Outlook

Ion mobility technology has achieved rapid growth in the past 2 decades, and commercial ion mobility platforms have emerged in an endless stream. IM-MS and its coupling with other analytical techniques have demonstrated outstanding advantages in greatly enhancing the confidence in the characterization and identification of chemical components, especially isomers, in different fields. In addition to experimental measurements and theoretical calculations, the prediction of CCS values can be more quickly and accurately achieved through ML methods, thereby establishing a dedicated multidimensional information database. Currently, some CCS databases based on different ML algorithms have been developed, such as MetCCS, LipidCCS, DeepCCS, and CCSbase. Moreover, the CCS databases have been applied in fields such as metabolomics, natural products, and foods. The development of quantum chemistry or molecular dynamics, such as the screening and optimization of 3D conformations and the determination of protonation/deprotonation sites, is helpful for obtaining the gas-phase structure closer to the measured state. The more comprehensive MD calculation methods can obtain more expected independent variables. The higher resolution of the IMS platform helps to obtain higher precision for dependent variable values. Last but not least, the newly developed appropriate feature screening approaches, ML or deep-learning algorithms, will help to greatly improve the accuracy of model fitting. With the further growth of IM-MS and the refinement of ML algorithms, it is believed that the prediction accuracy will be improved and that the database will be continuously expanded. The technology of predicting CCS values on the basis of using IM-MS and ML will also be deeply and widely used.

Author Contributions: Writing—original draft, X.L., H.W. and M.J.; investigation, X.L., H.W., M.D., X.X., B.X., Y.Z. and Y.Y.; funding acquisition, W.Y.; writing—review and editing, W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Tianjin Committee of Science and Technology of China (22ZYJDS00040) and the Science & Technology Program of the Haihe Laboratory of Modern Chinese Medicine (22HHZYJC00002).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable. No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors thank Ying Hu, Xue Li, Boxue Chen, and Feifei Yang for their contributions to editing and proofreading the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gabelica, V.; Shvartsburg, A.; Alexandre, A.; Afonso, C.; Barran, P.; Benesch, J.J.P.J.; Bleiholder, C.; Bowers, M.T.; Bilbao, A.; Bush, M.F.; et al. Recommendations for reporting ion mobility mass spectrometry measurements. *Mass Spectrom. Rev.* **2019**, *38*, 291–320. [\[CrossRef\]](#)
2. 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. [\[CrossRef\]](#)
3. Schrimpe-Rutledge, A.C.; Sherrod, S.D.; McLean, J.A. Improving the discovery of secondary metabolite natural products using ion mobility-mass spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 160–166. [\[CrossRef\]](#)
4. Moran-Garrido, M.; Camunas-Alberca, S.M.; Gil-de-la Fuente, A.; Mariscal, A.; Gradillas, A.; Barbas, C.; Sáiz, J. Recent developments in data acquisition, treatment and analysis with ion mobility-mass spectrometry for lipidomics. *Proteomics* **2022**, *22*, 2100328. [\[CrossRef\]](#)
5. Gabelica, V.; Marklund, E. Fundamentals of ion mobility spectrometry. *Curr. Opin. Chem. Biol.* **2018**, *42*, 51–59. [\[CrossRef\]](#)
6. Thomson, J.J.; Rutherford, E.X.L. On the passage of electricity through gases exposed to Röntgen rays. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1896**, *42*, 392–407. [\[CrossRef\]](#)
7. Ross, D.H.; Seguin, R.P.; Krinsky, A.M.; Xu, L.B. High-throughput measurement and machine learning-based prediction of collision cross sections for drugs and drug metabolites. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 1061–1072. [\[CrossRef\]](#)
8. Zuo, T.T.; Zhang, C.X.; Li, W.W.; Wang, H.D.; Hu, Y.; Yang, W.Z.; Jia, L.; Gao, X.M.; Guo, D. Offline two-dimensional liquid chromatography coupled with ion mobility-quadrupole time-of-flight mass spectrometry enabling four-dimensional separation and characterization of the multicomponents from white ginseng and red ginseng. *J. Pharm. Anal.* **2022**, *10*, 597–609. [\[CrossRef\]](#)
9. Liu, W.B.; Zhang, W.D.; Li, T.Z.; Zhou, Z.W.; Luo, M.D.; Chen, X.; Cai, Y.P.; Zhu, Z.J. Four-dimensional untargeted profiling of N-Acylethanolamine lipids in the mouse brain using ion mobility-mass spectrometry. *Anal. Chem.* **2022**, *94*, 12472–12480. [\[CrossRef\]](#)
10. 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. [\[CrossRef\]](#)
11. Ross, D.H.; Cho, J.H.; Xu, L.B. Breaking down structural diversity for comprehensive prediction of ion-neutral collision cross sections. *Anal. Chem.* **2020**, *92*, 4548–4557. [\[CrossRef\]](#)
12. 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.; et al. ISICLE: A quantum chemistry pipeline for establishing in silico collision cross section libraries. *Anal. Chem.* **2019**, *91*, 4346–4356. [\[CrossRef\]](#)
13. Ross, D.H.; Xu, L.B. Determination of drugs and drug metabolites by ion mobility-mass spectrometry: A review. *Anal. Chim. Acta* **2021**, *1154*, 338270. [\[CrossRef\]](#)
14. 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. [\[CrossRef\]](#)
15. Plante, P.L.; Francovic-Fontaine, É.; 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. [\[CrossRef\]](#)
16. Zhou, Z.W.; Luo, M.D.; Chen, X.; Yin, Y.D.; Xiong, X.; Wang, R.H.; Zhu, Z.J. Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nat. Commun.* **2020**, *11*, 4334. [\[CrossRef\]](#)
17. Koopman, J.; Grimme, S. From QCEIMS to QCxMS: A tool to routinely calculate CID mass spectra using molecular dynamics. *J. Am. Soc. Mass Spectr.* **2021**, *32*, 1735–1751. [\[CrossRef\]](#)
18. May, J.C.; McLean, J.A. Ion mobility-mass spectrometry: Time-dispersive instrumentation. *Anal. Chem.* **2015**, *87*, 1422–1436. [\[CrossRef\]](#)
19. Brinke, E.T.; Arrizabalaga-Larrañaga, A.; Blokland, M.H. Insights of ion mobility spectrometry and its application on food safety and authenticity: A review. *Anal. Chim. Acta* **2022**, *1222*, 340039. [\[CrossRef\]](#)
20. Zheng, X.Y.; Wojcik, R.; Zhang, X.; Ibrahim, Y.H.M.; Burnum-Johnson, K.E.; Orton, D.J.; Monroe, M.E.; Moore, R.J.; Smith, R.D.; Baker, E.S. Coupling front-end separations, ion mobility spectrometry, and mass spectrometry for enhanced multidimensional biological and environmental analyses. *Annu. Rev. Anal. Chem.* **2017**, *10*, 71–92. [\[CrossRef\]](#)
21. Mason, E.A.; Schamp, H.W., Jr. Mobility of gaseous ions in weak electric fields. *Ann. Phys.* **1958**, *4*, 233–270. [\[CrossRef\]](#)
22. Dodds, J.N.; May, J.C.; McLean, J.A. Investigation of the complete suite of the leucine and isoleucine isomers: Toward prediction of ion mobility separation capabilities. *Anal. Chem.* **2017**, *89*, 952–959. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Hinnenkamp, V.; Klein, J.; Meckelmann, S.W.; Balsaa, P.; Schmidt, T.C.; Schmitz, O.J. Comparison of CCS values determined by traveling wave ion mobility mass spectrometry and drift tube ion mobility mass spectrometry. *Anal. Chem.* **2018**, *90*, 12042–12050. [\[CrossRef\]](#)

24. Kemper, P.R.; Dupuis, N.F.; Bowers, M.T. A new, higher resolution, ion mobility mass spectrometer. *Int. J. Mass Spectrom.* **2009**, *287*, 46–57. [[CrossRef](#)]
25. Kirk, A.T.; Allers, M.; Cochems, P.; Langejuergen, J.; Zimmermann, S. A compact high resolution ion mobility spectrometer for fast trace gas analysis. *Analyst* **2013**, *138*, 5200–5207. [[CrossRef](#)] [[PubMed](#)]
26. Demelenne, A.; Nys, G.; Nix, C.; Fjeldsted, J.C.; Crommen, J.; Fillet, M. Separation of phosphorothioated oligonucleotide diastereomers using multiplexed drift tube ion mobility mass spectrometry. *Anal. Chim. Acta* **2022**, *1191*, 339297. [[CrossRef](#)] [[PubMed](#)]
27. Sipe, S.N.; Sanders, J.D.; Reinecke, T.; Clowers, B.H.; Brodbelt, J.S. Separation and collision cross section measurements of protein complexes afforded by a modular drift tube coupled to an orbitrap mass spectrometer. *Anal. Chem.* **2022**, *94*, 9434–9441. [[CrossRef](#)] [[PubMed](#)]
28. Sanders, J.D.; Shields, S.W.; Escobar, E.E.; Lanzillotti, M.B.; Butalewicz, J.P.; James, V.K.; Blevins, M.S.; Sipe, S.N.; Brodbelt, J.S. Enhanced ion mobility separation and characterization of isomeric phosphatidylcholines using absorption mode frontier transform multiplexing and ultraviolet photodissociation mass spectrometry. *Anal. Chem.* **2022**, *94*, 4252–4259. [[CrossRef](#)]
29. Lippmann, M.; Kirk, A.T.; Hitzemann, M.; Zimmermann, S. Compact and sensitive dual drift tube ion mobility spectrometer with a new dual field switching ion shutter for simultaneous detection of both ion polarities. *Anal. Chem.* **2020**, *92*, 11834–11841. [[CrossRef](#)]
30. George, A.C.; Schmitz-Afonso, I.; Marie, V.; Colsch, B.; Fenaille, F.; Afonso, C.; Loutelier-Bourhis, C. A re-calibration procedure for interoperable lipid collision cross section values measured by traveling wave ion mobility spectrometry. *Anal. Chem.* **2022**, *1226*, 340236. [[CrossRef](#)]
31. D’Atri, V.; Causon, T.; Hernandez-Alba, O.; Mutabazi, A.; Veuthey, J.L.; Cianferani, S.; Guilleme, D. Adding a new separation dimension to MS and LC-MS: What is the utility of ion mobility spectrometry? *J. Sep. Sci.* **2018**, *41*, 20–67. [[CrossRef](#)] [[PubMed](#)]
32. Ruotolo, B.T.; Benesch, J.L.P.; Sandercock, A.M.; Hyung, S.J.; Robinson, C.V. Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protoc.* **2008**, *3*, 1139–1152. [[CrossRef](#)] [[PubMed](#)]
33. Gelb, A.S.; Jarratt, R.E.; Huang, Y.; Dodds, E.D. A study of calibrant selection in measurement of carbohydrate and peptide ion-neutral collision cross sections by traveling wave ion mobility spectrometry. *Anal. Chem.* **2014**, *86*, 11396–11402. [[CrossRef](#)] [[PubMed](#)]
34. Bush, M.F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C.V.; Ruotolo, B.T. Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal. Chem.* **2010**, *82*, 9557–9565. [[CrossRef](#)] [[PubMed](#)]
35. Li, A.; Conant, C.R.; Zheng, X.; Bloodsworth, K.J.; Orton, D.J.; Garimella, S.V.B.; Attah, I.K.; Nagy, G.; Smith, R.D.; Ibrahim, Y.M. Assessing collision cross section calibration strategies for traveling wave-based ion mobility separations in structures for lossless ion manipulations. *Anal. Chem.* **2020**, *92*, 14976–14982. [[CrossRef](#)] [[PubMed](#)]
36. May, J.C.; Leaptrot, K.L.; Rose, B.S.; Moser, K.L.W.; Deng, L.L.; Maxon, L.; Debord, D.; McLean, J.A. Resolving power and collision cross section measurement accuracy of a prototype high-resolution ion mobility platform incorporating structures for lossless ion manipulation. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1126–1137. [[CrossRef](#)] [[PubMed](#)]
37. Giles, K.; Ujma, J.; Wildgoose, J.; Pringle, S.; Richardson, K.; Langridge, D.; Green, M. A cyclic ion mobility-mass spectrometry system. *Anal. Chem.* **2019**, *91*, 8564–8573. [[CrossRef](#)]
38. Ropartz, D.; Fanuel, M.; Ujma, J.; Palmer, M.; Giles, K.; Rogniaux, H. Structure determination of large isomeric oligosaccharides of natural origin through multipass and multistage cyclic traveling-wave ion mobility mass spectrometry. *Anal. Chem.* **2019**, *91*, 12030–12037. [[CrossRef](#)]
39. Colson, E.; Decroo, C.; Cooper-Shepherd, D.; Caulier, G.; Henoumont, C.; Laurent, S.; Winter, J.D.; Flammang, P.; Palmer, M.; Claereboudt, J.; et al. Discrimination of regioisomeric and stereoisomeric saponins from *Aesculus hippocastanum* seeds by ion mobility mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2228–2237. [[CrossRef](#)]
40. Rüger, C.P.; Le Maître, J.; Maillard, J.; Riches, E.; Palmer, M.; Afonso, C.; Giusti, P. Exploring complex mixtures by cyclic ion mobility high-resolution mass spectrometry: Application toward petroleum. *Anal. Chem.* **2021**, *93*, 5872–5881. [[CrossRef](#)]
41. Cavallero, G.J.; Zaia, J. Resolving heparan sulfate oligosaccharide positional isomers using hydrophilic interaction liquid chromatography-cyclic ion mobility mass spectrometry. *Anal. Chem.* **2022**, *94*, 2366–2374. [[CrossRef](#)] [[PubMed](#)]
42. Williamson, D.L.; Bergman, A.E.; Heider, E.C.; Nagy, G. Experimental measurements of relative mobility shifts resulting from isotopic substitutions with high-resolution cyclic ion mobility separations. *Anal. Chem.* **2022**, *94*, 2988–2995. [[CrossRef](#)] [[PubMed](#)]
43. Basit, A.; Pontis, S.; Piomelli, D.; Armirotti, A. Ion mobility mass spectrometry enhances low-abundance species detection in untargeted lipidomics. *Metabolomics* **2016**, *12*, 50. [[CrossRef](#)]
44. Luo, M.D.; Zhou, Z.W.; Zhu, Z.J. The application of ion mobility-mass spectrometry in untargeted metabolomics: From separation to identification. *J. Anal. Test.* **2020**, *4*, 163–174. [[CrossRef](#)]
45. Michelmann, K.; Silveira, J.A.; Ridgeway, M.E.; Park, M.A. Fundamentals of trapped ion mobility spectrometry. *J. Am. Soc. Mass Spectrom.* **2014**, *26*, 14–24. [[CrossRef](#)] [[PubMed](#)]
46. Silveira, J.A.; Ridgeway, M.E.; Park, M.A. High resolution trapped ion mobility spectrometry of peptides. *Anal. Chem.* **2014**, *86*, 5624–5627. [[CrossRef](#)]
47. Hernandez, D.R.; DeBord, J.D.; Ridgeway, M.E.; Kaplan, D.A.; Park, M.A.; Fernandez-Lima, F. Ion dynamics in a trapped ion mobility spectrometer. *Analyst* **2014**, *139*, 1913–1921. [[CrossRef](#)]

48. Vasilopoulou, C.G.; Sulek, K.; Brunner, A.D.; Meitei, N.S.; Schweiger-Hufnagel, U.; Meyer, S.W.; Barsch, A.; Mann, M.; Meier, F. Trapped ion mobility spectrometry and PASEF enable in-depth lipidomics from minimal sample amounts. *Nat. Commun.* **2020**, *11*, 331. [\[CrossRef\]](#)
49. Meier, F.; Brunner, A.D.; Koch, S.; Koch, H.; Lubeck, M.; Krause, M.; Goedecke, N.; Decker, J.; Kosinski, T.; Park, M.A.; et al. Online parallel accumulation-serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteom.* **2018**, *17*, 2534–2545. [\[CrossRef\]](#)
50. Charkow, J.; Rost, H.L. Trapped ion mobility spectrometry reduces spectral complexity in mass spectrometry-based proteomics. *Anal. Chem.* **2021**, *93*, 16751–16758. [\[CrossRef\]](#)
51. Adams, K.J.; Montero, D.; Aga, D.; Fernandez-Lima, F. Isomer separation of polybrominated diphenyl ether metabolites using nanoESI-TIM-MS. *Int. J. Ion Mobil. Spectrom.* **2016**, *19*, 69–76. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Meier, F.; Brunner, A.D.; Frank, M.; Ha, A.; Bludau, I.; Voytik, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; et al. diaPASEF: Parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods* **2020**, *17*, 1229–1236. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Helmer, P.O.; Nordhorn, I.D.; Korf, A.; Behrens, A.; Buchholz, R.; Zubeil, F.; Karst, U.; Hayen, H. Complementing matrix-assisted laser desorption/ionization-mass spectrometry imaging with chromatography data for improved assignment of isobaric and isomeric phospholipids utilizing trapped ion mobility-mass spectrometry. *Anal. Chem.* **2021**, *93*, 2135–2143. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Drakopoulou, S.K.; Damalas, D.E.; Baessmann, C.; Thomaidis, N.S. Trapped ion mobility incorporated in LC-HRMS workflows as an integral analytical platform of high sensitivity: Targeted and untargeted 4D-metabolomics in extra virgin olive oil. *J. Agri. Food Chem.* **2021**, *69*, 15728–15737. [\[CrossRef\]](#)
55. Zhang, J.D.; Kabir, K.M.M.; Lee, H.E.; Donald, W.A. Chiral recognition of amino acid enantiomers using high-definition differential ion mobility mass spectrometry. *Int. J. Mass Spectrom.* **2018**, *428*, 1–7. [\[CrossRef\]](#)
56. Dodds, J.N.; Baker, E.S. Ion mobility spectrometry: Fundamental concepts, instrumentation, applications, and the road ahead. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2185–2195. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Paglia, G.; Smith, A.J.; Astarita, G. Ion mobility mass spectrometry in the omics era: Challenges and opportunities for metabolomics and lipidomics. *Mass Spectrom. Rev.* **2022**, *41*, 722–765. [\[CrossRef\]](#)
58. Wang, H.D.; Wang, H.M.; Wang, X.Y.; Xu, X.Y.; Hu, Y.; Li, X.; Shi, X.J.; Wang, S.M.; Liu, J.; Qian, Y.X.; et al. A novel hybrid scan approach enabling the ion-mobility separation and the alternate data-dependent and data-independent acquisitions (HDDIDDA): Its combination with off-line two-dimensional liquid chromatography for comprehensively characterizing the multicomponents from Compound Danshen Dripping Pill. *Anal. Chim. Acta* **2022**, *1193*, 339320.
59. Picache, J.A.; Rose, B.S.; Balinski, A.; Leaptrot, K.L.; Sherrod, S.D.; May, J.C.; McLean, J.A. Collision cross section compendium to annotate and predict multi-omic compound identities. *Chem. Sci.* **2019**, *10*, 983–993. [\[CrossRef\]](#)
60. Wang, J.Y.; Yin, Y.H.; Zheng, J.Y.; Liu, L.F.; Yao, Z.P.; Xin, G.Z. Least absolute shrinkage and selection operator-based prediction of collision cross section values for ion mobility mass spectrometric analysis of lipids. *Analyst* **2022**, *147*, 1236–1244. [\[CrossRef\]](#)
61. Ahonen, L.; Fasciotti, M.; af Gennäs, G.B.; Kotiaho, T.; Daroda, R.J.; Eberlin, M.; Kostianen, R. Separation of steroid isomers by ion mobility mass spectrometry. *J. Chromatogr. A* **2013**, *1310*, 133–137. [\[CrossRef\]](#)
62. Wu, F.L.; Wu, X.S.; Chi, C.X.; Ding, C.F. Simultaneous differentiation of C = C position isomerism in fatty acids through ion mobility and theoretical calculations. *Anal. Chem.* **2022**, *94*, 12213–12220. [\[CrossRef\]](#)
63. Hofmann, J.; Hahm, H.S.; Seeberger, P.H.; Pagel, K. Identification of carbohydrate anomers using ion mobility-mass spectrometry. *Nature* **2015**, *526*, 241–244. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Ochoa, M.L.; Harrington, P.B. Detection of methamphetamine in the presence of nicotine using in situ chemical derivatization and ion mobility spectrometry. *Anal. Chem.* **2004**, *76*, 985–991. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Fenn, L.S.; McLean, J.A. Enhanced carbohydrate structural selectivity in ion mobility-mass spectrometry analyses by boronic acid derivatization. *Chem. Commun.* **2008**, *43*, 5505–5507. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Chouinard, C.D.; Cruzeiro, V.W.D.; Roitberg, A.E.; Yost, R.A. Experimental and theoretical investigation of sodiated multimers of steroid epimers with ion mobility-mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2016**, *28*, 323–331. [\[CrossRef\]](#)
67. Domalain, V.; Hubert-Roux, M.; Tognetti, V.; Joubert, L.; Lange, C.M.; Rouden, J.; Afonso, C. Enantiomeric differentiation of aromatic amino acids using traveling wave ion mobility-mass spectrometry. *Chem. Sci.* **2014**, *5*, 3234–3239. [\[CrossRef\]](#)
68. Asbury, G.R.; Hill, H.H. Using different drift gases to change separation factors (α) in ion mobility spectrometry. *Anal. Chem.* **2000**, *72*, 580–584. [\[CrossRef\]](#)
69. Dwivedi, P.; Wu, C.; Hill, H.H., Jr. Gas phase chiral separations by ion mobility spectrometry. *Anal. Chem.* **2006**, *78*, 8200–8206. [\[CrossRef\]](#)
70. Higton, D.; Palmer, M.E.; Vissers, J.P.C.; Mullin, L.G.; Plumb, R.S.; Wilson, I.D. Use of cyclic ion mobility spectrometry (cIM)-mass spectrometry to study the intramolecular transacylation of diclofenac acylglucuronide. *Anal. Chem.* **2021**, *93*, 7413–7421. [\[CrossRef\]](#)
71. Cooper-Shepherd, D.A.; Olivos, H.J.; Wu, Z.X.; Palmer, M.E. Exploiting self-association to evaluate enantiomeric composition by cyclic ion mobility-mass spectrometry. *Anal. Chem.* **2022**, *94*, 84418448. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Oganessian, I.; Hajduk, J.; Harrison, J.A.; Marchand, A.; Czar, M.F.; Zenobi, R. Exploring gas-phase MS methodologies for structural elucidation of branched N-Glycan isomers. *Anal. Chem.* **2022**, *94*, 10531–10539. [\[CrossRef\]](#) [\[PubMed\]](#)

73. May, J.C.; Knochenmuss, R.; Fjeldsted, J.C.; McLean, J.A. Resolution of isomeric mixtures in ion mobility using a combined demultiplexing and peak deconvolution technique. *Anal. Chem.* **2020**, *92*, 9482–9492. [[CrossRef](#)] [[PubMed](#)]
74. Chen, X.; Yin, Y.D.; Luo, M.D.; Zhou, Z.W.; Cai, Y.P.; Zhu, Z.J. Trapped ion mobility spectrometry-mass spectrometry improves the coverage and accuracy of four-dimensional untargeted lipidomics. *Anal. Chim. Acta* **2022**, *1210*, 339886. [[CrossRef](#)]
75. Kevin, P. Ion mobility-mass spectrometry of complex carbohydrates: Collision cross sections of sodiated *N*-linked glycans. *Anal. Chem.* **2013**, *85*, 5138–5145.
76. Leaptrot, K.L.; May, J.C.; Dodds, J.N.; McLean, J.A. Ion mobility conformational lipid atlas for high confidence lipidomics. *Nat. Commun.* **2019**, *10*, 985. [[CrossRef](#)] [[PubMed](#)]
77. Wojcik, R.; Nagy, G.; Attah, I.K.; Webb, I.K.; Garimella, S.V.B.; Weitz, K.K.; Hollerbach, A.; Monroe, M.E.; Ligare, M.R.; Nielson, F.F.; et al. SLIM ultrahigh resolution ion mobility spectrometry separations of isotopologues and isotopomers reveal mobility shifts due to mass distribution changes. *Anal. Chem.* **2019**, *91*, 11952–11962. [[CrossRef](#)]
78. Alves, T.O.; D'Almeida, C.T.S.; Victorio, V.C.M.; Souza, G.H.M.F.; Cameron, L.C.; Ferreira, M.S. Immunogenic and allergenic profile of wheat flours from different technological qualities revealed by ion mobility mass spectrometry. *J. Food Compos. Anal.* **2018**, *73*, 67–75. [[CrossRef](#)]
79. Paglia, G.; Angel, P.; Williams, J.P.; Richardson, K.; Olivos, H.J.; Thompson, J.W.; Menikarachchi, L.; Lai, S.; Walsh, C.; Moseley, A.; et al. Ion mobility-derived collision cross section as an additional measure for lipid fingerprinting and identification. *Anal. Chem.* **2015**, *87*, 1137–1144. [[CrossRef](#)]
80. Dwivedi, P.; Puzon, G.; Tam, M.; Langlais, D.; Jackson, S.; Kaplan, K.; Siems, W.F.; Schultz, A.J.; Xun, L.Y.; Woods, A.; et al. Metabolic profiling of *Escherichia coli* by ion mobility-mass spectrometry with MALDI ion source. *J. Mass Spectrom.* **2010**, *45*, 1383–1393. [[CrossRef](#)]
81. Djambazova, K.V.; Klein, D.R.; Migas, L.G.; Neumann, E.K.; Rivera, E.S.; Van de Plas, R.; Caprioli, R.M.; Spraggins, J.M. Resolving the complexity of spatial lipidomics using MALDI TIMS imaging mass spectrometry. *Anal. Chem.* **2020**, *92*, 13290–13297. [[CrossRef](#)] [[PubMed](#)]
82. Rainville, P.D.; Wilson, I.D.; Nicholson, J.K.; Isaac, G.; Mullin, L.; Langridge, J.I.; Plumb, R.S. Ion mobility spectrometry combined with ultra performance liquid chromatography/mass spectrometry for metabolic phenotyping of urine: Effects of column length, gradient duration and ion mobility spectrometry on metabolite detection. *Anal. Chim. Acta* **2017**, *982*, 1–8. [[CrossRef](#)] [[PubMed](#)]
83. Bennett, R.V.; Gamage, C.M.; Galhena, A.S.; Fernández, F.M. Contrast-enhanced differential mobility-desorption electrospray ionization-mass spectrometry imaging of biological tissues. *Anal. Chem.* **2014**, *86*, 3756–3763. [[CrossRef](#)] [[PubMed](#)]
84. Zhang, C.X.; Zuo, T.T.; Wang, X.Y.; Wang, H.D.; Hu, Y.; Li, Z.; Li, W.W.; Jia, L.; Qian, Y.X.; Yang, W.Z.; et al. Integration of data-dependent acquisition (DDA) and data-independent high-definition MSE (HDMSE) for the comprehensive profiling and characterization of multicomponents from *Panax japonicus* by UHPLC/IM-QTOF-MS. *Molecules* **2019**, *24*, 2708. [[CrossRef](#)] [[PubMed](#)]
85. Li, W.W.; Yang, X.N.; Chen, B.X.; Zhao, D.X.; Wang, H.D.; Sun, M.X.; Li, X.Y.; Liu, J.; Wang, S.M.; Mi, Y.G.; et al. Ultra-high performance liquid chromatography/ion mobility time-of-flight mass spectrometry-based untargeted metabolomics combined with quantitative assay unveiled the metabolic difference among the root, leaf, and flower bud of *Panax notoginseng*. *Arab. J. Chem.* **2021**, *1411*, 103409. [[CrossRef](#)]
86. Jeanne Dit Fouque, K.; Ramirez, C.E.; Lewis, R.L.; Koelmel, J.P.; Garrett, T.J.; Yost, R.A.; Fernandez-Lima, F. Effective liquid chromatography-trapped ion mobility spectrometry-mass spectrometry separation of isomeric lipid species. *Anal. Chem.* **2019**, *918*, 5021–5027. [[CrossRef](#)]
87. King, A.M.; Mullin, L.G.; Wilson, I.D.; Coen, M.; Rainville, P.D.; Plumb, R.S.; Gethings, L.A.; Maker, G.; Trengove, R. Development of a rapid profiling method for the analysis of polar analytes in urine using HILIC-MS and ion mobility enabled HILIC-MS. *Metabolomics* **2019**, *15*, 17. [[CrossRef](#)]
88. Szykuła, K.M.; Meurs, J.; Turner, M.A.; Creaser, C.S.; Reynolds, J.C. Combined hydrophilic interaction liquid chromatography-scanning field asymmetric waveform ion mobility spectrometry-time-of-flight mass spectrometry for untargeted metabolomics. *Anal. Bioanal. Chem.* **2019**, *411*, 6309–6317. [[CrossRef](#)]
89. Li, A.; Hines, K.M.; Xu, L.B. Lipidomics by HILIC-ion mobility-mass spectrometry. *Methods Mol. Biol.* **2020**, *2084*, 119–132.
90. Qian, Y.X.; Li, W.W.; Wang, H.M.; Hu, W.D.; Wang, H.D.; Zhao, D.X.; Hu, Y.; Li, X.; Gao, X.M.; Yang, W.Z. A four-dimensional separation approach by offline 2D-LC/IM-TOF-MS in combination with database-driven computational peak annotation facilitating the in-depth characterization of the multicomponents from *Atractylodes Macrocephalae* Rhizoma (*Atractylodes macrocephala*). *Arab. J. Chem.* **2021**, *142*, 102957.
91. Genangeli, M.; Heijens, A.M.M.; Rustichelli, A.; Schuit, N.D.; Micioni Di Bonaventura, M.V.M.D.; Cifani, C.; Vittori, S.; Siegel, T.P.; Heeren, R.M.A. MALDI-mass spectrometry imaging to investigate lipid and bile acid modifications caused by lentil extract used as a potential hypocholesterolemic treatment. *J. Am. Soc. Mass Spectrom.* **2019**, *3010*, 2041–2050. [[CrossRef](#)] [[PubMed](#)]
92. Shi, X.J.; Yang, W.Z.; Qiu, S.; Hou, J.J.; Wu, W.Y.; Guo, D. Systematic profiling and comparison of the lipidomes from *Panax ginseng*, *P. quinquefolius*, and *P. notoginseng* by ultrahigh performance supercritical fluid chromatography/high-resolution mass spectrometry and ion mobility-derived collision cross section measurement. *J. Chromatogr. A* **2018**, *1548*, 64–75. [[PubMed](#)]
93. Kurulugama, R.T.; Darland, E.; Kuhlmann, F.; Stafford, G.; Fjeldsted, J. Evaluation of drift gas selection in complex sample analyses using a high performance drift tube ion mobility-QTOF mass spectrometer. *Analyst* **2015**, *140*, 6834–6844. [[CrossRef](#)] [[PubMed](#)]

94. Paglia, G.; Williams, J.P.; Menikarachchi, L.; Thompson, J.W.; Tyldesley-Worster, R.; Halldórsson, S.; Rolfsson, O.; Moseley, A.; Grant, D.; Langridge, J.; et al. Ion mobility derived collision cross sections to support metabolomics applications. *Anal. Chem.* **2014**, *868*, 3985–3993. [[CrossRef](#)]
95. Zheng, X.Y.; Aly, N.A.; Zhou, Y.X.; Dupuis, K.T.; Bilbao, A.; Paurus, V.L.; Orton, D.J.; Wilson, R.; Payne, S.H.; Smith, R.D.; et al. A structural examination and collision cross section database for over 500 metabolites and xenobiotics using drift tube ion mobility spectrometry. *Chem. Sci.* **2017**, *811*, 7724–7736. [[CrossRef](#)] [[PubMed](#)]
96. Nichols, C.M.; Dodds, J.N.; Rose, B.S.; Picache, J.A.; Morris, C.B.; Codreanu, S.G.; May, J.C.; Sherrod, S.D.; McLean, J.A. Untargeted molecular discovery in primary metabolism: Collision cross section as a molecular descriptor in ion mobility-mass spectrometry. *Anal. Chem.* **2018**, *9024*, 14484–14492. [[CrossRef](#)]
97. Hernández-Mesa, M.; Bizec, B.L.; Monteau, F.; García-Campaña, A.M.; Dervilly-Pinel, G. Collision cross section CCS database: An additional measure to characterize steroids. *Anal. Chem.* **2018**, *907*, 4616–4625. [[CrossRef](#)]
98. Nye, L.C.; Williams, J.P.; Munjoma, N.C.; Letertre, M.P.M.; Coen, M.; Bouwmeester, R.; Martens, L.; Swann, J.R.; Nicholson, J.K.; Plumb, R.S.; et al. A comparison of collision cross section values obtained via travelling wave ion mobility-mass spectrometry and ultra high performance liquid chromatography-ion mobility-mass spectrometry: Application to the characterisation of metabolites in rat urine. *J. Chromatogr. A* **2019**, *1602*, 386–396. [[CrossRef](#)]
99. Poland, J.C.; Leaptrot, K.L.; Sherrod, S.D.; Flynn, C.R.; McLean, J.A. Collision cross section conformational analyses of bile acids via ion mobility-mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *318*, 1625–1631. [[CrossRef](#)]
100. May, J.C.; Goodwin, C.R.; Lareau, N.M.; Leaptrot, K.L.; Morris, C.B.; Kurulugama, R.T.; Mordehai, A.; Klein, C.; Barry, W.; Darland, E.; et al. Conformational ordering of biomolecules in the gas phase: Nitrogen collision cross sections measured on a prototype high resolution drift tube ion mobility-mass spectrometer. *Anal. Chem.* **2014**, *864*, 2107–2116. [[CrossRef](#)]
101. Stephan, S.; Hippler, J.; Köhler, T.; Deeb, A.A.; Schmidt, T.C.; Schmitz, O.J. Contaminant screening of wastewater with HPLC-IM-qTOF-MS and LC+ LC-IM-qTOF-MS using a CCS database. *Anal. Bioanal. Chem.* **2016**, *408*, 6545–6555. [[CrossRef](#)] [[PubMed](#)]
102. Hines, K.M.; Ross, D.H.; Davidson, K.L.; Bush, M.F.; Xu, L.B. Large-scale structural characterization of drug and drug-like compounds by high-throughput ion mobility-mass spectrometry. *Anal. Chem.* **2017**, *89*, 9023–9030. [[CrossRef](#)] [[PubMed](#)]
103. 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. C* **1996**, *100*, 16082–16086. [[CrossRef](#)]
104. Bleiholder, C.; Wyttenbach, T.; Bowers, M.T. A novel projection approximation algorithm for the fast and accurate computation of molecular collision cross sections (I). *Method. Int. J. Mass Spectrom.* **2011**, *308*, 1–10. [[CrossRef](#)]
105. 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. [[CrossRef](#)] [[PubMed](#)]
106. Hadavi, D.; Borzova, M.; Siegel, T.P.; Honing, M. Uncovering the behavior of ions in the gas-phase to predict the ion mobility separation of isomeric steroid compounds. *Anal. Chim. Acta* **2022**, *1200*, 339617. [[CrossRef](#)]
107. Przybylski, C.; Bonnet, V. Probing topology of supramolecular complexes between cyclodextrins and alkali metals by ion mobility-mass spectrometry. *Carbohydr. Polym.* **2022**, *297*, 120019. [[CrossRef](#)]
108. Turzo, S.B.A.; Seffernick, J.T.; Rolland, A.D.; Donor, M.T.; Heinze, S.; Prell, J.S.; Wysocki, V.H.; Lindert, S. Protein shape sampled by ion mobility mass spectrometry consistently improves protein structure prediction. *Nat. Commun.* **2022**, *13*, 4377. [[CrossRef](#)]
109. Song, X.C.; Dreolin, N.; Damiani, T.; Canellas, E.; Nerin, C. Prediction of collision cross section values: Application to non-intentionally added substance identification in food contact materials. *J. Agric. Food Chem.* **2022**, *70*, 1272–1281. [[CrossRef](#)]
110. Bijlsma, L.; Bade, R.; Celma, A.; Mullin, L.; Cleland, G.; Stead, S.; Hernandez, F.; Sancho, J.V. Prediction of collision cross-section values for small molecules: Application to pesticide residue analysis. *Anal. Chem.* **2017**, *89*, 6583–6589. [[CrossRef](#)]
111. Li, T.Z.; Yin, Y.D.; Zhou, Z.W.; Qiu, J.Q.; Liu, W.B.; Zhang, X.T.; He, K.W.; Zhu, Z.J. Ion mobility-based sterolomics reveals spatially and temporally distinctive sterol lipids in the mouse brain. *Nat. Commun.* **2021**, *12*, 4343. [[CrossRef](#)] [[PubMed](#)]
112. Stow, S.M.; Causon, T.J.; Zheng, X.; Kurulugama, R.T.; Mairinger, T.; May, J.C.; Rennie, E.E.; Baker, E.S.; Smith, R.D.; McLean, J.A.; et al. An interlaboratory evaluation of drift tube ion mobility-mass spectrometry collision cross section measurements. *Anal. Chem.* **2017**, *89*, 9048–9055. [[CrossRef](#)] [[PubMed](#)]
113. Hines, K.M.; May, J.C.; McLean, J.A.; Xu, L.B. Evaluation of collision cross section calibrants for structural analysis of lipids by traveling wave ion mobility-mass spectrometry. *Anal. Chem.* **2016**, *88*, 7329–7336. [[CrossRef](#)]
114. Paglia, G.; Astarita, G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.* **2017**, *12*, 797–813. [[CrossRef](#)] [[PubMed](#)]
115. Plachká, K.; Pezzatti, J.; Musenga, A.; Nicoli, R.; Kuuranne, T.; Rudaz, S.; Nováková, L.; Guilleme, D. Ion mobility-high resolution mass spectrometry in anti-doping analysis. Part I: Implementation of a screening method with the assessment of a library of substances prohibited in sports. *Anal. Chim. Acta* **2021**, *1152*, 338257. [[CrossRef](#)]
116. Jariyasopit, N.; Limjiasahapong, S.; Kurilung, A.; Sartyoungkul, S.; Wisanpitayakorn, P.; Nuntasaen, N.; Kuhakarn, C.; Reutrakul, V.; Kittakoop, P.; Sirivatanauskorn, Y.; et al. Traveling wave ion mobility-derived collision cross section database for plant specialized metabolites: An application to ventilaço harmandiana pierre. *J. Proteome Res.* **2022**, *21*, 2481–2492. [[CrossRef](#)] [[PubMed](#)]

117. Hernández-Mesa, M.; D'atri, V.; Barknowitz, G.; Fanuel, M.; Pezzatti, J.; Dreolin, N.; Ropartz, D.; Monteau, F.; Vigneau, E.; Rudaz, S.; et al. Interlaboratory and interplatform study of steroids collision cross section by traveling wave ion mobility spectrometry. *Anal. Chem.* **2020**, *92*, 5013–5022. [[CrossRef](#)]
118. Righetti, L.; Dreolin, N.; Celma, A.; McCullagh, M.; Barknowitz, G.; Sancho, J.V.; Dall'Asta, C. Travelling wave ion mobility-derived collision cross section for mycotoxins: Investigating interlaboratory and interplatform reproducibility. *J. Agric. Food Chem.* **2020**, *68*, 10937–10943. [[CrossRef](#)]
119. Hofmann, J.; Struwe, W.B.; Scarff, C.A.; Scrivens, J.H.; Harvey, D.J.; Pagel, K. Estimating collision cross sections of negatively charged *N*-glycans using traveling wave ion mobility-mass spectrometry. *Anal. Chem.* **2014**, *86*, 10789–10795. [[CrossRef](#)]
120. Das, S.; Tanemura, K.A.; Dinpazhoh, L.; Keng, M.; Schumm, C.; Leahy, L.; Asef, C.K.; Rainey, M.; Edison, A.S.; Fernández, F.M.; et al. In silico collision cross section calculations to aid metabolite annotation. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 750–759. [[CrossRef](#)]
121. Boschmans, J.; Jacobs, S.; Williams, J.P.; Palmer, M.; Richardson, K.; Giles, K.; Laphorn, C.; Herrebout, W.A.; Lemièrre, F.; Sobott, F. Combining density functional theory (DFT) and collision cross-section (CCS) calculations to analyze the gas-phase behavior of small molecules and their protonation site isomers. *Analyst* **2016**, *141*, 4044–4054. [[CrossRef](#)] [[PubMed](#)]
122. Ewing, S.A.; Donor, M.T.; Wilson, J.W.; Prell, J.S. Collidoscope: An improved tool for computing collisional cross-sections with the trajectory method. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 587–596. [[CrossRef](#)]
123. Bleiholder, C. A local collision probability approximation for predicting momentum transfer cross sections. *Analyst* **2015**, *140*, 6804–6813. [[CrossRef](#)] [[PubMed](#)]
124. 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. [[CrossRef](#)]
125. Larriba, C.; Hogan, J.C.J. Ion mobilities in diatomic gases: Measurement versus prediction with non-specular scattering models. *J. Phys. Chem. Lett.* **2013**, *117*, 3887–3901. [[CrossRef](#)]
126. Zanutto, L.; Heerdt, G.; Souza, P.C.T.; Araujo, G.; Araujo, M.S. High performance collision cross section calculation—HPCCS. *J. Comput. Chem.* **2018**, *39*, 1675–1681. [[CrossRef](#)]
127. Myers, C.A.; D'Esposito, R.J.; Fabris, D.; Ranganathan, S.V.; Chen, A.A. CoSIMS: An optimized trajectory-based collision simulator for ion mobility spectrometry. *J. Phys. Chem. B* **2019**, *123*, 4347–4357. [[CrossRef](#)]
128. Zhou, Z.W.; Xiong, X.; Zhu, Z.J. MetCCS predictor: A web server for predicting collision cross-section values of metabolites in ion mobility-mass spectrometry based metabolomics. *Bioinformatics* **2017**, *33*, 2235–2237. [[CrossRef](#)]
129. Lalli, P.M.; Corilo, Y.E.; Fasciotti, M.; Riccio, M.F.; Sa, G.F.; Daroda, R.J.; Souza, G.H.; McCullagh, M.; Bartberger, M.D.; Eberlin, M.N.; et al. Baseline resolution of isomers by traveling wave ion mobility mass spectrometry: Investigating the effects of polarizable drift gases and ionic charge distribution. *J. Mass Spectrom.* **2013**, *48*, 989–997. [[CrossRef](#)]
130. Lee, J.W.; Lee, H.H.L.; Davidson, K.L.; Bush, M.F.; Kim, H.I. Structural characterization of small molecular ions by ion mobility mass spectrometry in nitrogen drift gas: Improving the accuracy of trajectory method calculations. *Analyst* **2018**, *143*, 1786–1796. [[CrossRef](#)]
131. Shrivastav, V.; Nahin, M.; Hogan, C., J.; Larriba-Andaluz, C. Benchmark comparison for a multi-processing ion mobility calculator in the free molecular regime. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1540–1551. [[CrossRef](#)] [[PubMed](#)]
132. Marklund, E.G.; Degiacomi, M.T.; Robinson, C.V.; Baldwin, A.J.; Benesch, J.L.P. Collision cross sections for structural proteomics. *Structure* **2015**, *23*, 791–799. [[CrossRef](#)] [[PubMed](#)]
133. Belova, L.; Celma, A.; Van Haesendonck, G.; Lemièrre, F.; Sancho, J.V.; Covaci, A.; van Nuijs, A.L.N.; Bijlsma, L. Revealing the differences in collision cross section values of small organic molecules acquired by different instrumental designs and prediction models. *Anal. Chim. Acta* **2022**, *1229*, 340361. [[CrossRef](#)] [[PubMed](#)]
134. May, J.C.; McLean, J.A. Integrating ion mobility into comprehensive multidimensional metabolomics workflows: Critical considerations. *Metabolomics* **2022**, *18*, 104. [[CrossRef](#)] [[PubMed](#)]
135. Carracedo-Reboredo, P.; Liñares-Blanco, J.; Rodríguez-Fernández, N.; Cedrón, F.; Novoa, F.J.; Carballal, A.; Maojo, V.; Pazos, A.; Fernandez-Lozano, C. A review on machine learning approaches and trends in drug discovery. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 4538–4558. [[CrossRef](#)]
136. Mauri, A. alvaDesc: A tool to calculate and analyze molecular descriptors and fingerprints. In *Ecotoxicological QSARs*; Springer Nature: Berlin/Heidelberg, Germany, 2020; pp. 801–820.
137. Yang, F.; van Herwerden, D.; Preud'homme, H.; Samanipour, S. Collision cross section prediction with molecular fingerprint using machine learning. *Molecules* **2022**, *27*, 6424. [[CrossRef](#)] [[PubMed](#)]
138. Cereto-Massagué, A.; Ojeda, M.J.; Valls, C.; Mulero, M.; Garcia-Vallvé, S.; Pujadas, G. Molecular fingerprint similarity search in virtual screening. *Methods* **2015**, *71*, 58–63. [[CrossRef](#)]
139. Reymond, J.L.; Awale, M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.* **2012**, *3*, 649–657. [[CrossRef](#)]
140. Yap, C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem.* **2011**, *32*, 1466–1474. [[CrossRef](#)]
141. Ćimović, M.; Pezo, L.; Tešević, V.; Čabarkapa, I.; Todosijević, M. QSRR model for predicting retention indices of *Satureja kitaibelii* Wierzb. ex Heuff. essential oil composition. *Ind. Crops Prod.* **2020**, *154*, 112752. [[CrossRef](#)]

142. Soper-Hopper, M.T.; Petrov, A.S.; Howard, J.N.; Yu, S.S.; Forsythe, J.G.; Grover, M.A.; Fernández, F.M. Collision cross section predictions using 2-dimensional molecular descriptors. *Chem. Commun.* **2017**, *53*, 7624–7627. [[CrossRef](#)] [[PubMed](#)]
143. Soper-Hopper, M.T.; Vandegrift, J.; Baker, E.S.; Fernández, F.M. Metabolite collision cross section prediction without energy-minimized structures. *Analyst* **2020**, *145*, 5414–5418. [[CrossRef](#)] [[PubMed](#)]
144. Liu, A.L.; Venkatesh, R.; McBride, M.; Reichmanis, E.; Meredith, J.C.; Grover, M.A. Small data machine learning: Classification and prediction of poly (ethylene terephthalate) stabilizers using molecular descriptors. *ACS Appl. Polym. Mater.* **2020**, *2*, 5592–5601. [[CrossRef](#)]
145. Song, X.C.; Dreolin, N.; Canellas, E.; Goshawk, J.; Nerin, C. Prediction of collision cross-section values for extractables and leachables from plastic products. *Environ. Sci. Technol.* **2022**, *56*, 9463–9473. [[CrossRef](#)] [[PubMed](#)]
146. Moriwaki, H.; Tian, Y.S.; Kawashita, N.; Takagi, T. Mordred: A molecular descriptor calculator. *J. Cheminf.* **2018**, *10*, 4. [[CrossRef](#)] [[PubMed](#)]
147. Rainey, M.A.; Watson, C.A.; Asef, C.K.; Foster, M.R.; Baker, E.S.; Fernández, F.M. CCS Predictor 2.0: An open-source jupyter notebook tool for filtering out false positives in metabolomics. *Anal. Chem.* **2022**, *94*, 17456–17466. [[CrossRef](#)] [[PubMed](#)]
148. Cao, D.S.; Xu, Q.S.; Hu, Q.N.; Liang, Y.Z. ChemoPy: Freely available python package for computational biology and chemoinformatics. *Bioinformatics* **2013**, *29*, 1092–1094. [[CrossRef](#)] [[PubMed](#)]
149. Dong, J.; Cao, D.S.; Miao, H.Y.; Liu, S.; Deng, B.C.; Yun, Y.H.; Wang, N.N.; Lu, A.P.; Zeng, W.B.; Chen, A.F. ChemDes: An integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminf.* **2015**, *7*, 60. [[CrossRef](#)]
150. Du, J.X.; Chang, Y.Z.; Zhang, X.; Hu, C.Q. Development of a method of analysis for profiling of the impurities in phenoxymethylpenicillin potassium based on the analytical quality by design concept combined with the degradation mechanism of penicillins. *J. Pharm. Biomed. Anal.* **2020**, *186*, 113309. [[CrossRef](#)]
151. Willighagen, E.L.; Mayfield, J.W.; Alvarsson, J.; Berg, A.; Carlsson, L.; Jeliazkova, N.; Kuhn, S.; Pluskal, T.; Rojas-Chertó, M.; Spjuth, O.; et al. The Chemistry Development Kit (CDK) v2.0: Atom typing, depiction, molecular formulas, and substructure searching. *J. Cheminf.* **2017**, *9*, 33. [[CrossRef](#)]
152. Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): An open-source Java library for chemo- and bioinformatics. *J. Chem. Inf. Model.* **2003**, *43*, 493–500.
153. Broeckling, C.D.; Yao, L.; Isaac, G.; Gioioso, M.; Ianchis, V.; Vissers, J.P. Application of predicted collisional cross section to metabolome databases to probabilistically describe the current and future ion mobility mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 661–669. [[CrossRef](#)] [[PubMed](#)]
154. Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A.K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V.V.; Tanchuk, V.Y.; et al. Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.* **2011**, *25*, 533–554. [[CrossRef](#)]
155. Nielson, F.F.; Colby, S.M.; Thomas, D.G.; Renslow, R.S.; Metz, T.O. Exploring the impacts of conformer selection methods on ion mobility collision cross section predictions. *Anal. Chem.* **2021**, *93*, 3830–3838. [[CrossRef](#)]
156. Gonzales, G.B.; Smaghe, G.; Coelus, S.; Adriaenssens, D.; De Winter, K.; Desmet, T.; Raes, K.; Van Camp, J. Collision cross section prediction of deprotonated phenolics in a travelling-wave ion mobility spectrometer using molecular descriptors and chemometrics. *Anal. Chim. Acta* **2016**, *924*, 68–76. [[CrossRef](#)] [[PubMed](#)]
157. Ross, D.H.; Cho, J.H.; Zhang, R.T.; Hines, K.M.; Xu, L.B. LiPydomics: A python package for comprehensive prediction of lipid collision cross sections and retention times and analysis of ion mobility-mass spectrometry-based lipidomics data. *Anal. Chem.* **2020**, *92*, 14967–14975. [[CrossRef](#)] [[PubMed](#)]
158. Asef, C.K.; Rainey, M.A.; Garcia, B.M.; Gouveia, G.J.; Shaver, A.O.; Leach, F.E., III; Morse, A.M.; Edison, A.S.; McIntyre, L.M.; Fernández, F.M. Unknown metabolite identification using machine learning collision cross-section prediction and tandem mass spectrometry. *Anal. Chem.* **2023**, *95*, 1047–1056. [[CrossRef](#)] [[PubMed](#)]
159. 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.* **2019**, *92*, 1720–1729. [[CrossRef](#)]
160. Malik, A.; Saggi, M.K.; Rehman, S.; Sajjad, H.; Inyurt, S.; Bhatia, A.S.; Farooque, A.A.; Oudah, A.Y.; Yaseen, Z.M. Deep learning versus gradient boosting machine for pan evaporation prediction. *Eng. Appl. Comp. Fluid Mech.* **2022**, *16*, 570–587. [[CrossRef](#)]
161. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
162. Connolly, J.R.; Munoz-Muriedas, J.; Lapthorn, C.; Highton, D.; Vissers, J.P.; Webb, A.; Beaumont, C.; Dear, G.J. Investigation into small molecule isomeric glucuronide metabolite differentiation using *in silico* and experimental collision cross-section values. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1976–1986. [[CrossRef](#)]
163. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
164. Picache, J.A.; May, J.C.; McLean, J.A. Chemical class prediction of unknown biomolecules using ion mobility-mass spectrometry and machine learning: Supervised inference of feature taxonomy from ensemble randomization. *Anal. Chem.* **2020**, *92*, 10759–10767. [[CrossRef](#)] [[PubMed](#)]
165. Ieritano, C.; Lee, A.; Crouse, J.; Bowman, Z.; Mashmoushi, N.; Crossley, P.M.; Friebe, B.F.; Campbell, J.L.; Hopkins, W.S. Determining collision cross sections from differential ion mobility spectrometry. *Anal. Chem.* **2021**, *93*, 8937–8944. [[CrossRef](#)] [[PubMed](#)]

166. Song, X.C.; Canellas, E.; Dreolin, N.; Nerin, C.; Goshawk, J. Discovery and characterization of phenolic compounds in Bearberry (*Arctostaphylos uva-ursi*) leaves using liquid chromatography-ion mobility-high-resolution mass spectrometry. *J. Agric. Food Chem.* **2021**, *69*, 10856–10868. [[CrossRef](#)] [[PubMed](#)]
167. Wang, M.; Xu, X.Y.; Wang, H.D.; Wang, H.M.; Liu, M.Y.; Hu, W.D.; Chen, B.X.; Jiang, M.T.; Jing, Q.; Li, X.H.; et al. A multi-dimensional liquid chromatography/high-resolution mass spectrometry approach combined with computational data processing for the comprehensive characterization of the multicomponents from *Cuscuta chinensis*. *J. Chromatogr. A* **2022**, *1675*, 463162. [[CrossRef](#)]
168. Zhu, H.D.; Wu, X.D.; Huo, J.Y.; Hou, J.J.; Long, H.L.; Zhang, Z.J.; Wang, B.; Tian, M.H.; Chen, K.X.; Guo, D.A.; et al. A five-dimensional data collection strategy for multicomponent discovery and characterization in Traditional Chinese Medicine: *Gastrodia Rhizoma* as a case study. *J. Chromatogr. A* **2021**, *1653*, 462405. [[CrossRef](#)]
169. Yang, X.N.; Xiong, Y.; Wang, H.D.; Jiang, M.T.; Xu, X.Y.; Mi, Y.G.; Lou, J.; Li, X.H.; Sun, H.; Zhao, Y.Y.; et al. Multicomponent characterization of the flower bud of *Panax notoginseng* and its metabolites in rat plasma by ultra-high performance liquid chromatography/ion mobility quadrupole time-of-flight mass spectrometry. *Molecules* **2022**, *27*, 9049. [[CrossRef](#)]
170. Mullin, L.; Jobst, K.; DiLorenzo, R.A.; Plumb, R.; Reiner, E.J.; Yeung, L.W.; Jogsten, I.E. Liquid chromatography-ion mobility-high resolution mass spectrometry for analysis of pollutants in indoor dust: Identification and predictive capabilities. *Anal. Chem. Acta* **2020**, *1125*, 29–40. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.