Pacific Northwest
NATIONAL LABORATORY

PNNL-34731

A Joint Modeling/Experimental Approach to Characterize Ionization and Fragmentation of SOA Molecules with CIMS

August 2023

Edoardo Aprà
Amity Andersen
Daniel Mejia-Rodriguez

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from
the Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062
www.osti.gov
ph: (865) 576-8401
fox: (865) 576-5728
email: reports@osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
or (703) 605-6000
email: info@ntis.gov
Online ordering: http://www.ntis.gov

PNNL-34731

A Joint Modeling/Experimental Approach to Characterize Ionization and Fragmentation of SOA Molecules with CIMS

August 2023

Edoardo Aprà
Amity Andersen
Daniel Mejia-Rodriguez

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Abstract

The Molecular Dynamics technique Collective variable hyperdynamics (CVHD) interfaced with density functional tight-binding at the GFN1-xTB and GFN2-xTB levels of electronic structure theory has been applied to the H^+GAG positively charged peptide system to explore the fragmentation of this simple tripeptide under relatively low temperature/energy conditions. Unlike conventional chemical dynamics simulations which can only be performed up to hundreds of picoseconds when coupled with semi-empirical Hamiltonians, CVHD is able to capture the long-time dynamics of multiple proton hopping, the formation of reversible ring structures, and ester rearrangement of the H^+GAG system prior to fragmentation. The CVHD method applied to H^+GAG is also able to uncover alternative fragmentation pathways not considered previously such as glycine fragmentation on the N-terminal side of the charged peptide and new cyclic cationic species. These findings in the H^+GAG case have implications for other peptide systems.

Summary

NWChem has been interfaced with both the PLUMED code for performing metadynamics and hyperdynamics as well as the TBlite code to bring tight-binding style semi-empirical electronic structure theory to the NWChem suite of codes. The PLUMED code with the Collective variable hyperdynamics (CVHD) module from Bal et al. was interfaced to NWChem and CP2K to perform CVHD simulations on the H^+GAG tripeptide system and on the singly protonated caffeine molecule to showcase this capability for mass spectrometry applications at relatively low energies/temperatures. The CVHD simulations were done at the GFN1-xTB and GFN2-xTB levels of tight-binding theory from the TBlite code. Density functional theory (DFT) CVHD simulations were also tried, but the DFT level of electronic structure theory proved to be too computationally demanding for the tripeptide system, which required at least 40,000 steps (100 picoseconds) of simulation time for meaningful results. The tight-binding results showed different fragmentation patterns for the GFN1-xTB and GFN2-xTB levels of theory. Interestingly, it is the GFN1-xTB older-generation tight-binding theory that correctly predicts the glycine plus $[b_2]^+$ (and isomers of $[b_2]^+$) entropically-favored fragmentation of H^+GAG as the major fragmentation path. Moreover, a second competitive glycine plus isomers of $[b_2]^+$ reaction pathway involving an ester intermediate formed from the N-terminal end of the peptide was found at the GFN1-xTB level of theory. These two reaction pathways are only minor paths at the GFN2-xTB level of theory. For GFN2-xTB, the H_2O plus $[b_3]^+$ non-ring isomers is the prevalent fragmentation reaction path. It is conjectured that differences in proton binding to the various O and N atom sites and van der Waals interaction influence the discrepancy between the GFN1-xTB and GFN2-xTB levels of theory.

Acknowledgments

This research was supported by the **M/Q Initiative**, under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

Abstract ii
Summary iii
Acknowledgments iv
1.0 Introduction 3
2.0 Computational Methods 5
2.1 CREST optimization of initial structures prior to dynamics 5
2.2 300 K equilibration of structures 5
2.3 Collective-variable-driven hyperdynamics simulations 5
2.4 Nudged-elastic band description of novel fragmentation paths 7
3.0 Results & Discussion 8
3.1 Protonation state issue, dynamic proton transfer 8
3.2 Primary mechanism of Glycine elimination 8
3.3 New alternative mechanism of glycine elimination 10
3.4 Minor competing paths such as dehydration and ammonia elimination 11
3.5 Rates of reaction for selected reactions 12
3.6 Caffeine 13
4.0 Conclusions 16
References 17

Figures

Figure 1. Proton hopping mechanisms. 8
Figure 2. Most common fragmentation mechanism. 9
Figure 3. Nudged-elastic bands along the most common fragmentation mechanism. 10
Figure 4. New mechanism of glycine elimination where glycine eliminated from the N-terminal end of the peptide (rather than the C-terminal end in the standard reaction pathway) through an “ester”-like intermediate. The charged fragment is an isomer of the $[b_2]^+$ fragment of the original glycine elimination mechanism 11
Figure 5. New mechanism of glycine elimination. 11
Figure 6. Alternate H^+GAG fragmentation reaction path leading to $[b_3]^+$ and H_2O 12
Figure 7. Most stable protomer of caffeine. 13
Figure 8. Only charged fragment obtained in CVHD simulations at 1000 K. 14
Figure 9. Five different fragments with m/z = 138 obtained in CVHD simulations at 1500 K and 2000 K. 14

Tables

Table 1. Global distortion bond parameters (lengths in Angstroms) used during the collective-variable-driven hyperdynamics simulations. 6

Table 2. Comparison of reaction rates using a total of sixty-four trajectories.....12

1.0 Introduction

Accurate determination of small molecule and biological polymer fragmentation processes in mass spectrometry (MS) and tandem mass spectrometry (MS/MS, MS^n) is the desired outcome for the characterization of chemical compounds in samples containing multiple unknowns. Prior to the determination of unknowns in sample mixtures, the fragmentation of known samples with single target compounds will need to be studied well to provide a comprehensive database of fragmentation patterns for search queries. Even for the simplest compounds, this can be a challenge due to multiple experimental variables including sample preparation, ionization method, mass spectrometer apparatus setup, and collision gas, to name a few experimental conditions. (Szabo and Janaky 2015; Hufsky, Scheubert, and Böcker 2014; Cui, Lu, and Lee 2018) Informatics efforts using machine learning and artificial intelligence (ML/AI) techniques are promising data science methods to tackle this problem and potentially fill in gaps in knowledge. (Lavecchia 2019; Baum et al. 2021) However, these techniques typically have not included chemical physics information in the building of models.

A first step in including chemical physics information is to perform chemical dynamics simulations from first principles on many compounds to provide training material for subsequent ML/AI efforts. Quantum chemistry tools have been invaluable for elucidating mechanisms of mass spectrometry fragmentation (Mookherjee, 2019,2022) Traditionally, quantum chemistry tools alone require experimental knowledge of fragmentation products and chemical intuition to determine plausible pathways for MS fragmentation. These tools are often coupled with statistical rate theory methods such as Rice–Ramsperger–Kassel–Marcus (RRKM) analysis to give further insight into the kinetic processes and branching ratios.

Chemical dynamics methods pioneered by Hase and coworkers (Homayoon, 2018; Martin, 2019; Macaluso, 2019; Pratihar, 2017; Spezia, 2009; Gu, 2020; Spezia, 2016; Malik, 2019; Malik, 2020) can more directly ascertain the reaction pathways for MS fragmentation (thermally or via collision-induced dissociation (CID)) and even discover new pathways overlooked by the application of chemical intuition. These studies involve on-the-fly molecular dynamics (MD) with a quantum chemistry engine (e.g., semi-empirical Hamiltonian, density functional theory (DFT)) interfaced with a molecular dynamics driver (e.g. VENUS (Hu, 1991; Lourderaj, 2014) interfaced with the NWChem (Aprà, 2020) DFT module). More recently, Koopman and Grimme have developed an automated tool for calculating CID mass spectra using molecular dynamics coupled with their xTB code (QCxMS) (Koopman, 2021). Conventional chemical dynamics simulations of these previous works are limited in time scale (tens to hundreds of picoseconds if semi-empirical Hamiltonian methods are used) (Koopman, 2021; Martin, 2019). For low temperatures and low energy collisions, conventional chemical dynamics may not result in reactions observed in MS operating at these conditions (Martin, 2019).

In this work, we propose the use of hyperdynamics to accelerate the chemical dynamics at low temperatures/energies where intramolecular vibrational energy redistribution (IVR) is likely. Like conventional chemical dynamics, hyperdynamics can be used to explore the potential energy surface for unexpected alternate mechanisms as well as determine kinetics and branch ratios. Moreover, chemical dynamics, in general, can provide a history of intermediates along reaction pathways, which gives important mechanistic insight and patterns that can be used to train neural networks. Since hyperdynamics allows one to explore long time scale dynamics (possibly up to seconds (Bal, 2015)), exploration of the peptide “mobile proton model” (Dongre,1996) (a

relatively slow process) is possible as well as subsequent fragmentation resulting from proton transfer.

There are a few hyperdynamics schemes available for the acceleration of molecular dynamics and chemical dynamics simulations. Initial hyperdynamics schemes were pioneered by Voter and coworkers (Voter, 1997). The first schemes involved temperature biasing to accelerate molecular dynamics simulations. Later hyperdynamics schemes involved bias potentials to boost molecular dynamics simulations (Huang, 2015). Of these methods, the Bond Boost potential designed by Miron and Fichthorn, where the bias is zeroed for any geometry that is beyond acritical length between nearby atoms (Miron, 2003) is an appealing intuitive form of hyperdynamics for the study of gas phase ion fragmentation where bond vibration and breaking would be the main drivers of fragmentation. Later, Bal and Neyts, inspired by the Bond Boost hyperdynamics implementation of Miron and Fichthorn used the functional collective variable (CV) forms of Tiwary and van de Walle (Tiwary, 2013) and metadynamics concepts to design a "self-learning" collective variable hyperdynamics (CVHD) method (Bal, 2015). We employed this generalized implementation of hyperdynamics in our study of gas phase ion fragmentation in mass spectrometry.

Here, we focus on selected protonated amino acids and very small peptides containing a few amino acid residues. Amino acids, peptides, and proteins are present in all living systems. Thus, a quantitative assessment of mass spectrometry fragmentation patterns of these compounds is important in biological (Biemann, 1992) biomedical (Srebalus, 2007), pharmaceutical (Geoghegan, 2005; Srebalus, 2007; Ackermann, 2008) food (Mamone, 2009; Kaufmann, 2012; Lu, 2018) natural products (Bouslimani, 2014; Cheng, 2010; Bode, 2012) Agricultural (Vega, 2005; Marbaix, 2016) environmental (Lebedev, 2013; Richardson, 2008; Mcilvin, 2021; Zhou, 2020) and forensic (Valletta, 2021; Van Steendam, 2013; Hoffmann, 2015) analyses.

We will present new insight into the mass spectrometry observations of the H^+GAG peptide ion recently published by Mookherjee and Armentrout (Mookherjee, 2019). In their work, they present standard quantum chemistry calculations, density functional theory (DFT) and Møller-Plesset second-order perturbation theory (MP2), of stationary points based on their knowledge of the ion products from guided-ion mass spectrometry experiments. They did not perform chemical dynamics simulations on the H^+GAG peptide ion system. Our study is a first-of-a-kind study of the H^+GAG peptide ion system using CVHD chemical dynamics, and we report a novel alternate low-energy pathway that can compete with the primary lowest energy pathway discovered by Mookherjee and Armentrout via standard quantum chemistry calculations. Moreover, CVHD chemical dynamics can capture the proton mobility expected from long dynamics.

2.0 Computational Methods

2.1 CREST optimization of initial structures prior to dynamics

The conformer-rotamer ensemble sampling tool (CREST) (Grimme, 2019; Pracht, 2020; Pracht, 2021) was used to screen an initial set of structure candidates that was later refined using DFT. CREST searches for the most stable conformers using a combination of metadynamics sampling and genetic z-matrix crossing. CREST efficiency is rooted in the use of the semi-empirical extended tight binding (xTB) methodology (Grimme, 2016; Grimme, 2017; Bannwarth, 2019; Bannwarth, 2021). The xTB methods were developed to circumvent the generality issues from other popular tight-binding methodologies, with a primary aim to describe geometries, (vibrational) frequencies, and non-covalent interactions (GFN). The first-generation GFN1-xTB method (Grimme, 2017) employs a very similar Hamiltonian as the third-generation Density Functional Tight-Binding (DFTB3) approach, including the use of the D3(BJ) empirical dispersion correction. However, the GFN1-xTB Hamiltonian does not rely on atom pairwise parameters. GFN1-xTB was readily adopted in some quantum-chemical software packages, including CP2K. Some deficiencies of GFN1-xTB were solved with the second-generation GFN2-xTB method (Bannwarth, 2019). GFN2-xTB goes beyond the monopole-type atomic pairwise electrostatic interaction, includes anisotropic exchange-correlation contributions, and uses a charge-dependent (D4) empirical dispersion correction. Here, we used the GFN1-xTB implementation available in CP2K, as well as the GFN2-xTB method from a new interface between NWChem and the TBlite library.

2.2 300 K equilibration of structures

The most important protomers and conformers found by CREST were equilibrated at 300 K in order to generate a rich set of starting points for subsequent analysis. To this end, unbiased NVT MD simulations were performed using GFN1-xTB (CP2K), GFN2-xTB (NWChem), or the Perdew-Burke-Ernzerhoff exchange-correlation functional (NWChem).

In order to speed up Kohn-Sham Density Functional Theory calculations, the auxiliary density functional theory (ADFT) method was also implemented in NWChem. ADFT uses the density-fitting approach to reduce the computational effort involved in obtaining the classical Coulomb interaction, while also taking advantage of the same fitted density in order to lower the effort associated with the numerical integration of the exchange-correlation contribution. Depending on the calculation details (orbital and fitting basis sets, numerical integration grid density, screening thresholds), ADFT might achieve speed-ups of one order of magnitude as compared to a standard DFT calculation. The downside of the ADFT approach is that it offers no advantage when combined with traditional metaGGA exchange-correlation functionals since both the orbital and fitting basis sets must be computed on the numerical integration grid. However, ADFT does offer speedups with the “deorbitalized” versions of metaGGAs. The overall accuracy of the ADFT methodology depends on the quality of the fitting basis set. Almost identical relative energies are attainable if a good fitting basis is used.

2.3 Collective-variable-driven hyperdynamics simulations

Collective-variable-driven hyperdynamics (CVHD) (Bal, 2015) is a combination of conventional metadynamics and hyperdynamics. Metadynamics is an enhanced sampling method useful for the calculation of free energies and accelerating rare events (Laio, 2002). In metadynamics, free-energy minima are filled by a history-dependent potential term, allowing the exploration of

the free-energy surface as a function of a few collective variables. Hyperdynamics is a similar technique in the sense that a bias potential is added in order to accelerate the dynamics of a given system. However, hyperdynamic simulations remove the bias potential at transition states, resulting in correct relative dynamics but losing free energy information (Voter, 1997). The goal of a CVHD simulation is to bias the distortion χ_i of any given collective variable s_i. If the collective variables are chosen properly, a large distortion χ should correspond with a bond-breaking event. The CVHD method was implemented in a fork of the PLUMED library (Bussi, 2019) maintained by Kristof Bal.

MD simulations were obtained with either CP2K or NWChem interfaced to the PLUMED library with the CVHD module enabled. The NWChem-PLUMED interface is available in the development version of NWChem, and the CP2K-PLUMED interface is available in the release version. *NVT* dynamics were obtained using the canonical sampling through velocity rescaling (CVSR) thermostat set at 1000 K, a 0.25 fs time-step, and removing only center-of-mass translations. Extended TB trajectories were collected for 40,000 steps (100 picoseconds).

During these simulations, bond-breaking events were biased using a 0.25 kcal/mol metadynamics-type hill added every one hundred time steps and using a bias factor of 20. Bonds were grouped in "Global Distortion" collective variables η following the standard CVHD procedure. Local distortions were computed using the parameters shown in Table 1 according to

$$\chi_i = \begin{cases} 0 & \text{if } r_i \leqslant r_{\text{ref}} \\ \frac{r_i - r_{\text{ref}}}{r_{\text{ref}}} & \text{if } r_i > r_{\text{ref}} \end{cases}$$

Eq. 1

from where the global distortion η can be obtained as

$$\eta = \left(\sum_i \chi_i^p \right)^{1/p}$$

Eq. 2

At the beginning of a CVHD run, a list of *real* bonds with $r_{min} \leq r \leq r_{max}$ is generated. This list is updated for every bond-breaking event detected by the CVHD module according to the bond parameters provided in Table 1.

Table 1. Global distortion bond parameters (lengths in Angstroms) used during the collective-variable-driven hyperdynamics simulations.

Bond	r_{min}	r_{ref}	r_{max}	p
C-C	1.40	1.55	2.00	6
C-N	0.90	1.40	1.90	6
C-O	1.00	1.30	1.90	6

C-H	0.60	1.10	1.60	6
O-H	0.50	1.00	1.50	6
N-H	0.50	1.00	1.50	6

In the CVHD method, the amount the clock is advanced, Δt_b depends on the strength of the bias at the current position of the trajectory. When the bias potential, $\Delta V_b(r)$, is zero, Δt_b equals the time step size of an unbiased MD simulation, Δt_{MD}. The average improvement in CVHD compared to a direct MD simulation is given by the so-called boost factor, τ

$$\tau = \frac{t_b}{t_{MD}} = \langle e^{\beta \Delta V_b(\mathbf{r})} \rangle \quad \text{Eq. 3}$$

where $\beta = 1/kT$. Thus, the time sampled in a CVHD run (also known as hypertime) is just $t_b = \tau\, t_{MD}$.

2.4 Nudged-elastic band description of novel fragmentation paths

The initial fragmentation pathways were generated using the tight-binding xTB-GFN2 methodology using the TBlite library linked to NWChem. Elementary reactions were first approximated by constrained optimizations and then refined by using the Nudged-elastic band (NEB) method (Jónsson, 1998) with 15 beads per reaction, a spring force constant of 10^{-3} Ha/Bohr^2, and the limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newton optimizer. For the initial NEB band, the reactant and product were superimposed prior to the linear interpolation step.

NEB calculations were run until the maximum force component was smaller than 10^{-3} Ha/Bohr. Possible intermediates were further extracted whenever the NEB calculation produced a band with oscillatory behavior. The accuracy setting for xTB-GFN2 was set to 0.01.

Once the xTB-GFN2 fragmentation pathway was completed, local minima were reoptimized using the PBE-D3(BJ) density functional approximation. All surviving minima were then connected using the NEB methodology with the same parameters as described above. Again, possible intermediates were obtained from bands with oscillatory behavior.

The fragmentation pathways obtained with PBE-D3(BJ) were used as the initial guess for the r^2 SCAN-D3(BJ) and M06-2X density functional approximations.

All DFT calculations used the SIMINT molecular integral package, with a Schwarz-screening threshold of 10^{-12} a.u. PBE-D3(BJ) calculations used a medium numerical integration grid, while r^2 SCAN-D3(BJ) and M06-2X used a fine grid.

3.0 Results & Discussion

3.1 Protonation state issue, dynamic proton transfer

In all trajectories, we observe fast, accelerated proton hopping among N and O bearing groups (as depicted in Figure 1). This observation gives credence to the “mobile proton model” proposed by Dongre et al. (Dongre, 1996), and its role in giving rise to various fragmentation products.

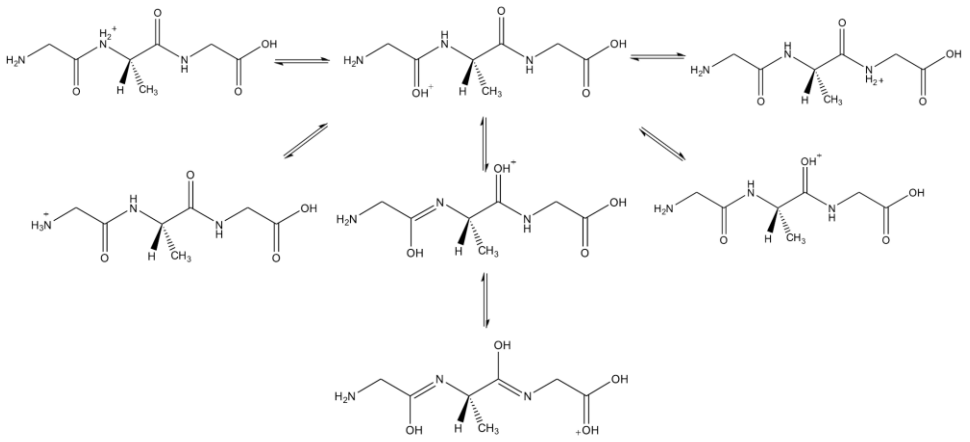

Figure 1. Proton hopping mechanisms.

3.2 Primary mechanism of Glycine elimination

As detailed by Armentrout and Mookherjee (Mookherjee, 2019), the lowest energy fragmentation process arises from the loss of neutral glycine to form the protonated 2-(aminomethyl)-4-methyl oxazolone ion (H^+AMMOx, $[b_2]^+$). The lowest energy path according to Armentrout and Mookherjee’s experimental and theoretical results is the $[b_2]^+$ plus glycine path shown in Figure 2.

We found, overall, a very similar path as the one described for the protonated triglycine system (Mookherjee, 2017) for all the tested methods, including xTB-GFN1, xTB-GFN2, PBE-D3, r^2SCAN-D3, and M06-2X. However, there are some important differences between the lower-level methodologies and the DFT ones, as shown in Figure 3.

Following the nomenclature in Refs. (Mookherjee, 2017) and (Mookherjee, 2019), the xTB methods predict that the [O^{1t}]-ctgttgtt conformer is the global minimum. In contrast, the [O^{1t}]-ctgttttt conformer is the global minimum with all DFT methods.

Both xTB methodologies predict more compact paths, as measured by the total distance moved by each atom, than their DFT counterparts. The xTB-GFN1 method predicts the lowest reaction barriers, below all other methodologies by at least 5 kcal/mol. Moreover, the xTB-GFN1 method predicts that the sum of the energies of the separated fragments is just 28 kcal/mol above the global minimum. In comparison, all other methodologies predict that the energy of the separated fragments is, at least, 42 kcal/mol above their respective global minimum.

Overall, the tight-binding methodologies preferred a cyclic oxazolone intermediate before fragmentation, whereas the DFT ones predicted that fragmentation and cyclization occurred in a concerted fashion.

All three density functional approximations evaluated here produced rather similar paths.

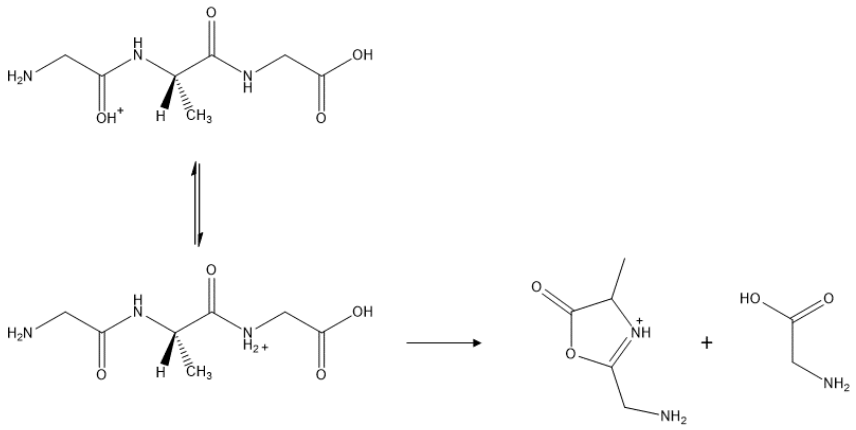

Figure 2. Most common fragmentation mechanism.

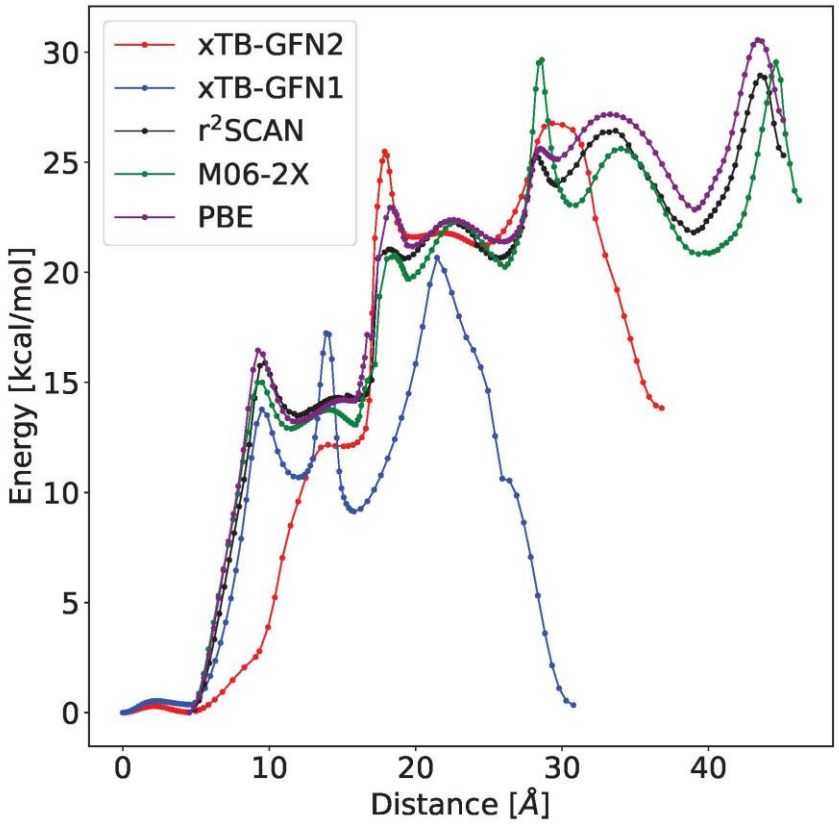

Figure 3. Nudged-elastic bands along the most common fragmentation mechanism.

Unfortunately, the DFT trajectory runs are too expensive for getting meaningful hyperdynamics trajectories (i.e., only short runs were possible and often did not lead to fragmentation). Thus, we proceeded with only GFN1-xTB and GFN2-xTB runs. Interestingly, when comparing the CVHD trajectories of GFN1-xTB versus GFN2-xTB runs, the latter semi-empirical xTB method gives the trajectories where the $[b_2]^+$ (or isomers thereof) plus glycine are, correctly, the dominant fragmentation products. As we discuss in the next section, the GFN1-xTB runs also show a competitive alternative fragmentation leading to glycine and a product with the same m/z as the $[b_2]^+$ ion. Since the GFN2-xTB model is meant to be an improvement on the older GFN1-xTB model, this serves as a cautionary tale that more development is needed into semi-empirical methods to improve their reliability in simulating the dynamics of model small peptide systems such as H^+GAG.

3.3 New alternative mechanism of glycine elimination

A new alternative path for glycine elimination was obtained during the CVHD simulations with both tight-binding methodologies (see Figure 4). As we mentioned above, the GFN1-xTB method shows that this alternative path is competitive with the direct lowest energy and entropically favored path to $[b_2]^+$ and glycine. We further explored the zero-temperature limit of this path by using the nudged-elastic band method.

Major differences between the xTB-GFN2 method and DFT were obtained in this case, as seen in Figure 5. While the highest xTB-GFN2 barrier is about 30 kcal/mol above the global minimum, DFT ones are around 50 kcal/mol. The tight-binding path required more atomic motion than the DFT-based ones in order to locate all the stable local minima.

The energy of the fragmented products is about 55 kcal/mol above the global minimum with any methodology.

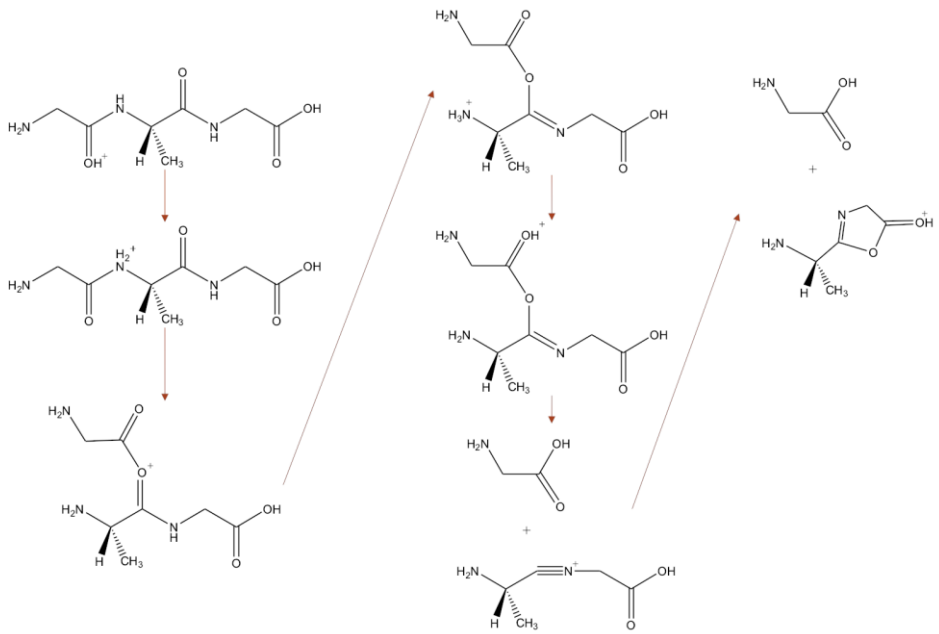

Figure 4. New mechanism of glycine elimination where glycine eliminated from the N-terminal end of the peptide (rather than the C-terminal end in the standard reaction pathway) through an “ester”-like intermediate. The charged fragment is an isomer of the $[b_2]^+$ fragment of the original glycine elimination mechanism.

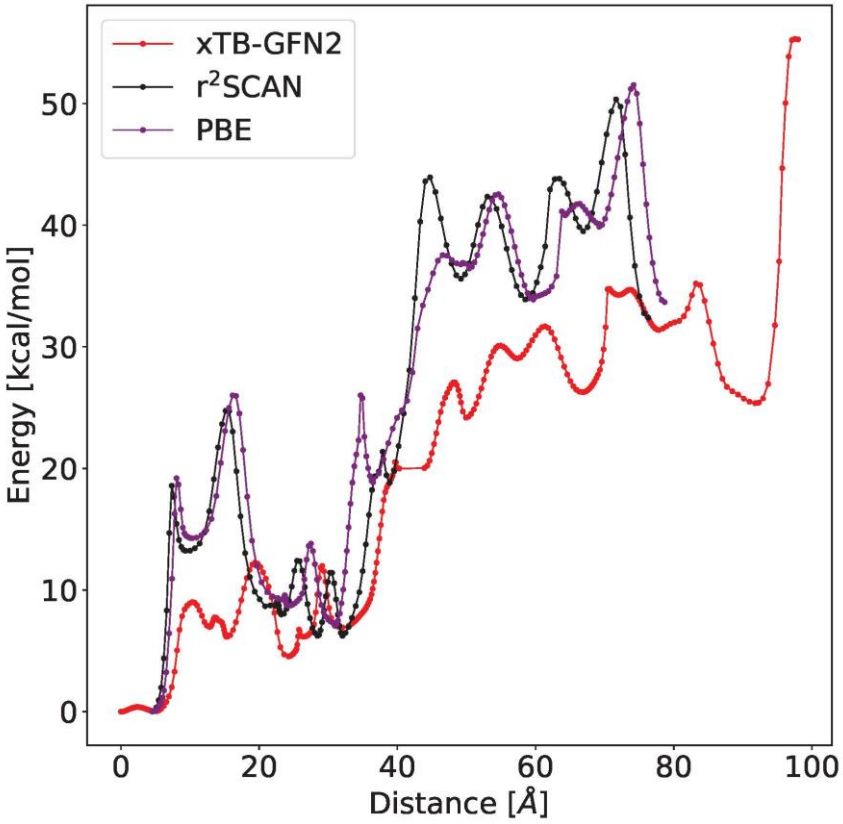

Figure 5. New mechanism of glycine elimination.

3.4 Minor competing paths such as dehydration and ammonia elimination

As mentioned previously, the GFN1-xTB CVHD H^+GAG trajectories primarily lead to glycine and $[b_2]^+$ (or isomers thereof) fragmentation (this fragmentation pattern is minor for GFN2-xTB). For GFN1-xTB, dehydration of $HGAG^+$ (H_2O and isomers of $[b_3]^+$, including $[b_3]^+$ itself) is one possible minor pathway for $HGAG^+$ fragmentation (see Figure 6). H^+GAG fragmentation leading to H_2O and isomers of $[b_3]^+$ (including $[b_3]^+$ itself) is the major path according to the GFN2-xTB CVHD trajectories, though it is not the entropically favored fragmentation path. The glycine elimination and the dehydration paths are very competitive, and any discrepancy in the accuracy in the electronic structure theory is shown by the dynamics leading to these two groups of product pathways.

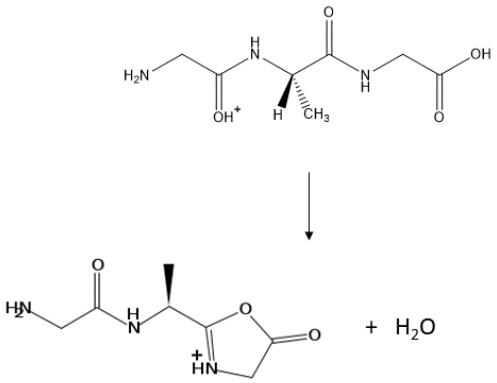

Figure 6. Alternate H^+GAG fragmentation reaction path leading to $[b_3]^+$ and H_2O.

Another minor set of pathways leads to the formation of ammonia plus products.

3.5 Rates of reaction for selected reactions

The CVHD methods allows for estimation of the thermal rate of reaction through the calculation of hypertime according to the Eq. 3 boost factor. These rates can be compared with rates from statistical rate theories such as RRKM. However, time constraints did not allow for the calculation of RRKM rate constants for comparison. Also, due to time constraints, we only report a fraction of the 100+ trajectory data we have analyzed thus far for GFN1-xTB and GFN2-xTB runs (see Table 2). Currently, finding the hypertime of the rate-limiting step of these reactions is a time-consuming task involving sight inspection of output files and calculations by hand. However, we have been writing Python scripts to automate this hypertime calculation task, and we are in the process of testing these scripts.

Here we only report the rate estimate orders of magnitude for the selected fragmentation reactions of a set of 64 trajectories simulated at the GFN1-xTB level of theory. Only 13 out of the 64 trajectories had fragmentation occur in the 100 picoseconds simulation time. The boost factor at the time of fragmentation was multiplied by the MD simulation time at fragmentation to compute the hypertime rate of fragmentation. For the glycine+$[b_2]^+$ isomers reactions, the alternative path is very competitive with the standard path (as mentioned earlier). The rate estimate range for these two path groups is quite wide. The glycine+$[b_2]^+$ isomers reactions can occur faster than the H_2O+$[b_3]^+$ isomer paths; however, the rate estimate ranges do overlap. Thus the H_2O+$[b_3]^+$ isomer fragmentation reaction can be competitive with the glycine+$[b_2]^+$ isomers pathways.

Table 2. Comparison of reaction rates using a total of sixty-four trajectories.

Reaction Type	Number of trajectories	Rate Estimate (s^{-1} at 1000 K)
Standard glycine+$[b_2]^+$/ $[b_2]^+$ isomers	6	10^7-10^9
Alternative glycine+$[b_2]^+$/ $[b_2]^+$ isomers	4	10^8-10^{11}

H_2O+$[b_3]^+$/ $[b_3]^+$ isomers	2	10^8
NH_3+product	1	10^6

3.6 Caffeine

The CVHD approach can also be used to simulate high-energy CIMS spectra. We benchmarked this capability using the semiempirical xTB-GFN2 via the NWChem interface to TBlite. To this end, we sampled a 200 picoseconds unbiased trajectory at 300 K started with the most stable protomer of Caffeine shown in Figure 7.

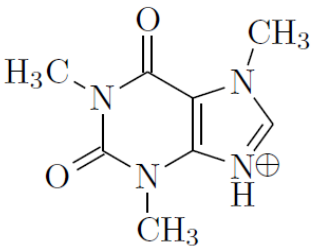

Figure 7. Most stable protomer of caffeine.

We discarded the first 50 picoseconds, and extracted 500 structures from the remaining 150 picoseconds. These structures were equally divided and randomly assigned to the 1000 K, 1500 K, 2000 K, 2500 K, and 3000 K temperature groups to obtain CIMS fragmentation patterns at different mimicking different collision energies. Bias was added to C-C, C-O, C-N, C-H, N-H, and O-H bonds using the parameters presented in Table 1.

CVHD trajectories were run for 50 picoseconds with a time step of 10 au (0.24188 fs) using the Nosé-Hoover chain thermostat with a frequency of 3000 cm^{-1}. The electronic temperature used in xTB-GFN2 matched the ionic temperature. Center-of-mass translations were removed during the simulation, but rotations were kept, as this might add a centrifugal force that aids in fragmentation mechanisms.

At 1000 K, only one trajectory (out of 100) resulted in fragmentation with a neutral methyl isocyanate leaving group (Figure 8). At 1500 K, half of the trajectories ended up with the same methyl isocyanate neutral loss but ended up with five different fragments with m/z = 138 (Figure 9). Almost all trajectories showed fragmentation at 2000 K, with most of them leading to the same m/z = 138 isomers shown in Figure 9, but about 10% of the trajectories led to charged products with lower masses, including the methenium cation. Trajectories generated at 2500 K and 3000 K showed excessive fragmentation as compared to experimental CIMS spectra, most of the time ending up in methenium cations. The excessive fragmentation is a result of excess energy provided both by the thermal bath as well as by the bias introduced in the CVHD method. An alternative approach to obtain meaningful results at such high temperatures is to run shorter MD simulations to avoid the excessive buildup of metadynamics bias. The simulated CIMS spectrum shown in Figure 10 was generated using 5 picoseconds trajectories at 2500 K. This spectrum shows that the CVHD method allows the prediction of CIMS spectra with

reasonable agreements to other computational methods but using very short trajectories. Therefore, the CVHD method should enable non-negligible computational savings.

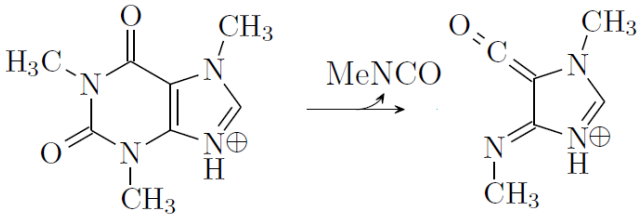

Figure 8. Only charged fragment obtained in CVHD simulations at 1000 K.

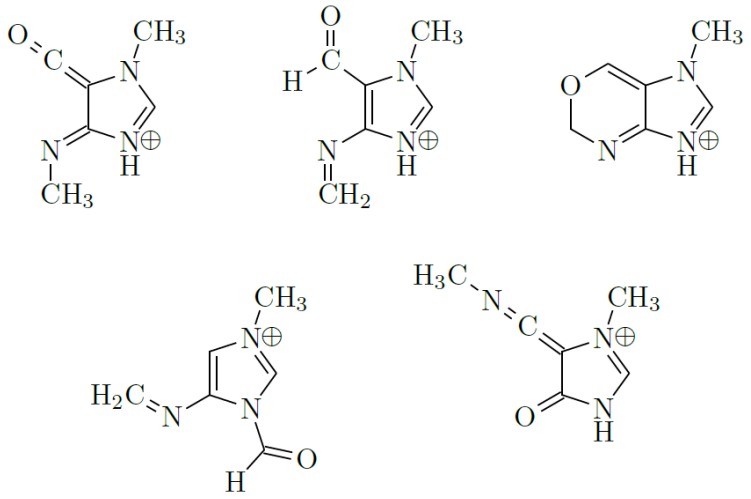

Figure 9. Five different fragments with m/z = 138 obtained in CVHD simulations at 1500 K and 2000 K.

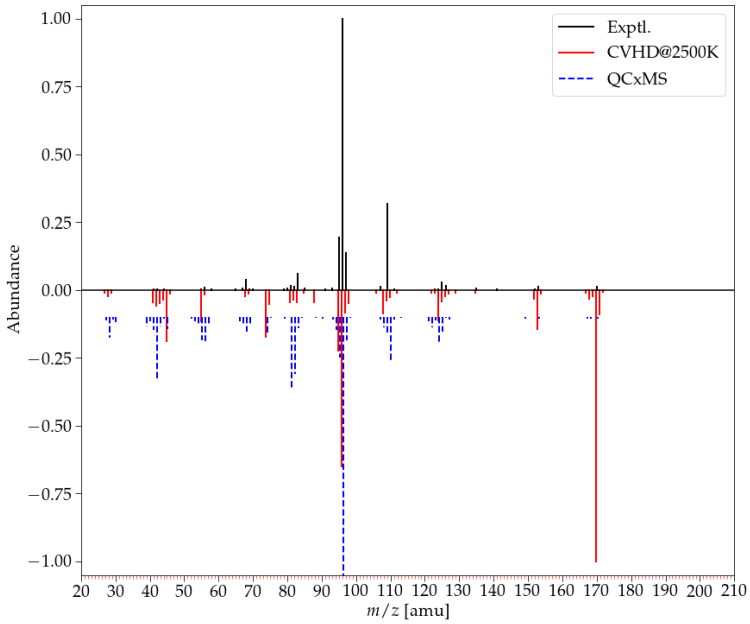

Figure 10. Simulated Caffeine CIMS spectra obtained with CVHD approach and the xTB-GFN2 semiempirical method. Comparison to other experimental and theoretical references is also shown.

4.0 Conclusions

NWChem has been interfaced to PLUMED (metadynamics and hyperdynamics simulations) and TBlite (xTB extended tight-binding semiempirical quantum chemistry method). With these two interfaces, one can now run collective-variable hyperdynamics simulations at the first-principles and semiempirical tight-binding levels of theory. The interfaces with NWChem also allows users to run metadynamics simulations with PLUMED with first-principles and semi-empirical quantum chemistry levels of theory. In addition to these developments with the NWChem code, we have been implementing a few SLURM, shell, and Python scripts to automate the process of farming multiple instances of CVHD simulations and on-the-fly analysis of the resulting simulations data. These scripts are still in the process of testing. We applied the new NWChem interfaces to PLUMED and TBlite to protonated small peptide (namely H^+GAG) and small molecule (caffeine) systems of interest to mass spectrometry. For the H^+GAG system, we demonstrate fast hopping of protons with the accelerated molecular dynamics, which would be slow processes with conventional molecular dynamics. We see that this proton hopping influences the resulting fragmentation product yields of the H^+GAG system. At the GFN1-xTB level of tight-bind theory, we find that the CVHD simulation fragmentation favors the production of glycine and $[b_2]^+$ ions (and isomers thereof) in line with experimental and first-principles stationary-point calculations favoring this entropically-favorable product profile. While the more direct reaction path for production of glycine at C-terminal of the H^+GAG tripeptide is the more favorable fragmentation path, we find a competitive alternative reaction path for production of glycine at from the N-terminal end of the tripeptide, which involves an ester-like intermediate. At the GFN2-xTB level of theory, the fragmentation of H^+GAG to H_2O and $[b_3]^+$ (and isomers thereof) is much more favorable compared to fragmentation to glycine and $[b_2]^+$ ions (and isomers thereof), which is counter to experimental and theoretical observations of Mookeherjee and Armentrout (GFN2-xTB is the newest generation xTB level of theory). It is possible that there is a fundamental difference the way the proton affinities and van der Waals interactions between GFN1-xTB and GFN2-xTB are handled, which needs to be explored further and could lead to improvements in the xTB method. Fragmentation to NH_3 and products is also an important minor fragmentation path for H^+GAG. We also report some preliminary estimates on the reaction rates derived from the hypertime calculations for a subset of trajectories we have performed thus far.

References

Szabo, Zoltan, and Tamas Janaky. 2015. “Challenges and developments in protein identification using mass spectrometry.” TrAC Trends in Analytical Chemistry 69:76–87.

Hufsky, Franziska, Kerstin Scheubert, and Sebastian Böcker. 2014. "Computational mass spectrometry for small-molecule fragmentation." *TrAC Trends in Analytical Chemistry* 53:41-48.

Cui, Liang, Haitao Lu, and Yie Hou Lee. 2018. "Challenges and emergent solutions for LC-MS/MS based untargeted metabolomics in diseases." *Mass spectrometry reviews* 37 (6):772-792.

Lavecchia, Antonio. 2019. "Deep learning in drug discovery: opportunities, challenges and future prospects." *Drug discovery today* 24 (10):2017-2032.

Baum, Zachary J., Xiang Yu, Philippe Y. Ayala, Yanan Zhao, Steven P. Watkins, and Qiongqiong Zhou. 2021. "Artificial intelligence in chemistry: current trends and future directions." *Journal of Chemical Information and Modeling* 61 (7):3197-3212.

Mookherjee, A., and P. B. Armentrout. 2019. "Thermodynamics and reaction mechanisms for decomposition of a simple protonated tripeptide, H+ GAG: a guided ion beam and computational study." *Journal of The American Society for Mass Spectrometry* 30 (6):1013-1027.

Mookherjee, Abhigya, and P. B. Armentrout. 2022. "Thermodynamics and Reaction Mechanisms for Decomposition of a Simple Protonated Tripeptide, H+ GGA: From H^+ GGG to H^+ GAG to H^+ GGA." *Journal of the American Society for Mass Spectrometry* 33 (2):355-368.

Homayoon, Zahra, Veronica Macaluso, Ana Martin-Somer, Maria Carolina Nicola Barbosa Muniz, Itamar Borges, William L. Hase, and Riccardo Spezia. 2018. "Chemical dynamics simulations of CID of peptide ions: comparisons between TIK (H+) 2 and TLK (H+) 2 fragmentation dynamics, and with thermal simulations." *Physical Chemistry Chemical Physics* 20 (5):3614-3629.

Martin Somer, Ana, Veronica Macaluso, George L. Barnes, Li Yang, Subha Pratihar, Kihyung Song, William L. Hase, and Riccardo Spezia. 2019. "Role of chemical dynamics simulations in mass spectrometry studies of collision-induced dissociation and collisions of biological ions with organic surfaces." *Journal of the American Society for Mass Spectrometry* 31 (1):2-24.

Macaluso, Veronica, Debora Scuderi, Maria Elisa Crestoni, Simonetta Fornarini, Davide Corinti, Enzo Dalloz, Emilio Martinez-Nunez, William L. Hase, and Riccardo Spezia. 2019. "L-Cysteine modified by S-sulfation: consequence on fragmentation processes elucidated by tandem mass spectrometry and chemical dynamics simulations." *The Journal of Physical Chemistry A* 123 (17):3685-3696.

Pratihar, Subha, Xinyou Ma, Zahra Homayoon, George L. Barnes, and William L. Hase. 2017. "Direct chemical dynamics simulations." *Journal of the American Chemical Society* 139 (10):3570-3590.

Spezia, Riccardo, Jean-Yves Salpin, Marie-Pierre Gaigeot, William L. Hase, and Kihyung Song. 2009 "Protonated urea collision-induced dissociation. Comparison of experiments and chemical dynamics simulations." *The Journal of Physical Chemistry A* 113 (50):13853-13862.

Gu, Meng, Jiaxu Zhang, William L. Hase, and Li Yang. 2020. "Direct dynamics simulations of the thermal fragmentation of a protonated peptide containing arginine." *ACS omega* 5 (3):1463-1471.

Spezia, Riccardo, Ana Martin-Somer, Veronica Macaluso, Zahra Homayoon, Subha Pratihar, and William L. Hase. 2016. "Unimolecular dissociation of peptides: statistical vs. non-statistical fragmentation mechanisms and time scales." *Faraday Discussions* 195:599-618.

Malik, Abdul, Yu-Fu Lin, Subha Pratihar, Laurence A. Angel, and William L. Hase. 2019. "Direct dynamics simulations of fragmentation of a Zn (II)-2Cys-2His oligopeptide. Comparison with mass spectrometry collision-induced dissociation." *The Journal of Physical Chemistry A* 123 (32):6868-6885.

Malik, Abdul, Laurence A. Angel, Riccardo Spezia, and William L. Hase. 2020. "Collisional dynamics simulations revealing fragmentation properties of Zn (II)-bound poly-peptide." *Physical Chemistry Chemical Physics* 22 (26):14551-14559.

Hu, Xiche, William L. Hase, and Tony Pirraglia. 1991. "Vectorization of the general Monte Carlo classical trajectory program VENUS." *Journal of computational chemistry* 12 (8):1014-1024.

Lourderaj, Upakarasamy, Rui Sun, Swapnil C. Kohale, George L. Barnes, Wibe A. De Jong, Theresa L. Windus, and William L. Hase. 2014. "The VENUS/NWChem software package. Tight coupling between chemical dynamics simulations and electronic structure theory." *Computer Physics Communications* 185 (3):1074-1080.

Aprà, Edoardo, Eric J. Bylaska, Wibe A. De Jong, Niranjan Govind, Karol Kowalski, Tjerk P. Straatsma, Marat Valiev et al. 2020. "NWChem: Past, present, and future." *The Journal of chemical physics* 152(18):184102.

Koopman, Jeroen, and Stefan Grimme. 2021. "From QCEIMS to QCxMS: A tool to routinely calculate CID mass spectra using molecular dynamics." *Journal of the American Society for Mass Spectrometry* 32 (7):1735-1751.

Bal, Kristof M., and Erik C. Neyts. 2015. "Merging metadynamics into hyperdynamics: accelerated molecular simulations reaching time scales from microseconds to seconds." *Journal of chemical theory and computation* 11 (10):4545-4554.

Dongré, Ashok R., Jennifer L. Jones, Árpád Somogyi, and Vicki H. Wysocki. 1996. "Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: evidence for the mobile proton model." *Journal of the American Chemical Society* 118 (35):8365-8374.

Voter, Arthur F. 1997. "Hyperdynamics: Accelerated molecular dynamics of infrequent events." *Physical Review Letters* 78 (20):3908.

Huang, Chen, Danny Perez, and Arthur F. Voter. 2015. "Hyperdynamics boost factor achievable with an ideal bias potential." *The Journal of chemical physics* 143 (7):074113.

Miron, Radu A., and Kristen A. Fichthorn. 2003. "Accelerated molecular dynamics with the bond-boost method." *The Journal of chemical physics* 119 (12):6210-6216.

Tiwary, Pratyush, and Axel van de Walle. 2013. "Accelerated molecular dynamics through stochastic iterations and collective variable-based basin identification." *Physical Review B* 87 (9):094304.

Biemann, K. 1992. "Mass spectrometry of peptides and proteins." *Annual review of biochemistry* 61 (1):977-1010.

Srebalus Barnes, Catherine A., and Amareth Lim. "Applications of mass spectrometry for the structural characterization of recombinant protein pharmaceuticals." *Mass spectrometry reviews* 26(3 (2007): 370-388.

Geoghegan, Kieran F., and Michele A. Kelly. "Biochemical applications of mass spectrometry in pharmaceutical drug discovery." *Mass spectrometry reviews* 24(3 (2005): 347-366.

Ackermann, Bradley L., Michael J. Berna, James A. Eckstein, Lee W. Ott, and Ajai K. Chaudhary. 2008. "Current applications of liquid chromatography/mass spectrometry in pharmaceutical discovery after a decade of innovation." *Annu. Rev. Anal. Chem.* 1: 357-396.

Mamone, Gianfranco, Gianluca Picariello, Simonetta Caira, Francesco Addeo, and Pasquale Ferranti. 2009. "Analysis of food proteins and peptides by mass spectrometry-based techniques." *Journal of Chromatography A* 1216 (43):7130-7142.

Kaufmann, Anton. 2012. "The current role of high-resolution mass spectrometry in food analysis." *Analytical and bioanalytical chemistry* 403 (5):1233-1249.

Lu, Haiyan, Hua Zhang, Konstantin Chingin, Jianliang Xiong, Xiaowei Fang, and Huanwen Chen. 2018. "Ambient mass spectrometry for food science and industry." *TrAC Trends in Analytical Chemistry* 107:99-115.

Bouslimani, Amina, Laura M. Sanchez, Neha Garg, and Pieter C. Dorrestein. 2014. "Mass spectrometry of natural products: current, emerging and future technologies." *Natural product reports* 31 (6):718-729.

Cheng, Ka-Wing, Chi-Chun Wong, Mingfu Wang, Qing-Yu He, and Feng Chen. 2010. "Identification and characterization of molecular targets of natural products by mass spectrometry." *Mass Spectrometry Reviews* 29 (1):126-155.

Bode, Helge B., Daniela Reimer, Sebastian W. Fuchs, Ferdinand Kirchner, Christina Dauth, Carsten Kegler, Wolfram Lorenzen, Alexander O. Brachmann, and Peter Grün. 2012. "Determination of the absolute configuration of peptide natural products by using stable isotope labeling and mass spectrometry." *Chemistry–A European Journal* 18 (8):2342-2348.

Vega, A. Belmonte, A. Garrido Frenich, and JL Martínez Vidal. 2005. "Monitoring of pesticides in agricultural water and soil samples from Andalusia by liquid chromatography coupled to mass spectrometry." *Analytica Chimica Acta* 538 (1-2):117-127.

Marbaix, Helene, Dimitri Budinger, Marc Dieu, Olivier Fumiere, Nathalie Gillard, Philippe Delahaut, Sergio Mauro, and Martine Raes. 2016. "Identification of proteins and peptide biomarkers for detecting banned processed animal proteins (PAPs) in meat and bone meal by mass spectrometry." *Journal of agricultural and food chemistry* 64 (11):2405-2414.

Lebedev, Albert T. 2013. "Environmental mass spectrometry." *Annual review of analytical chemistry* 6:163-189.

Richardson, Susan D. 2008. "Environmental mass spectrometry: emerging contaminants and current issues." *Analytical chemistry* 80 (12): 4373-4402.

McIlvin, Matthew R., and Mak A. Saito. 2021. "Online Nanoflow Two-Dimension Comprehensive Active Modulation Reversed Phase–Reversed Phase Liquid Chromatography High-Resolution Mass Spectrometry for Metaproteomics of Environmental and Microbiome Samples." *Journal of proteome research* 20 (9):4589-4597.

Zhou, Wei, Weiqi Xu, Hwajin Kim, Qi Zhang, Pingqing Fu, Douglas R. Worsnop, and Yele Sun. 2020. "A review of aerosol chemistry in Asia: insights from aerosol mass spectrometer measurements." *Environmental Science: Processes & Impacts* 22 (8):1616-1653.

Valletta, Mariangela, Sara Ragucci, Nicola Landi, Antimo Di Maro, Paolo Vincenzo Pedone, Rosita Russo, and Angela Chambery. 2021. "Mass spectrometry-based protein and peptide profiling for food frauds, traceability and authenticity assessment." *Food Chemistry* 365:130456.

Van Steendam, Katleen, Marlies De Ceuleneer, Maarten Dhaenens, David Van Hoofstat, and Dieter Deforce. 2013. "Mass spectrometry-based proteomics as a tool to identify biological matrices in forensic science." *International journal of legal medicine* 127:287-298.

Hoffmann, William D., and Glen P. Jackson. 2015. "Forensic mass spectrometry." *Annual Review of Analytical Chemistry* 8:419-440.

Grimme, Stefan. 2019. "Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations." *Journal of chemical theory and computation* 15 (5):2847-2862.

Pracht, Philipp, Fabian Bohle, and Stefan Grimme. 2020. "Automated exploration of the low-energy chemical space with fast quantum chemical methods." *Physical Chemistry Chemical Physics* 22 (14):7169-7192.

Pracht, Philipp, and Stefan Grimme. 2021. "Calculation of absolute molecular entropies and heat capacities made simple." *Chemical science* 12 (19):6551-6568.

Grimme, Stefan, and Christoph Bannwarth. 2016. "Ultra-fast computation of electronic spectra for large systems by tight-binding based simplified Tamm-Dancoff approximation (sTDA-xTB)." *The Journal of chemical physics* 145 (5):054103.

Grimme, Stefan, Christoph Bannwarth, and Philip Shushkov. 2017. "A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements (Z= 1–86)." *Journal of chemical theory and computation* 13 (5):1989-2009.

Pracht, Philipp, Eike Caldeweyher, Sebastian Ehlert, and Stefan Grimme. 2019. "A robust non-self-consistent tight-binding quantum chemistry method for large molecules."

Bannwarth, Christoph, Eike Caldeweyher, Sebastian Ehlert, Andreas Hansen, Philipp Pracht, Jakob Seibert, Sebastian Spicher, and Stefan Grimme. 2021. "Extended tight-binding quantum chemistry methods." *Wiley Interdisciplinary Reviews: Computational Molecular Science* 11 (2):e1493.

Laio, Alessandro, and Michele Parrinello. 2002. "Escaping free-energy minima." *Proceedings of the national academy of sciences* 99 (20):12562-12566.

Bussi, Giovanni, and Gareth A. Tribello. 2019. "Analyzing and biasing simulations with PLUMED." *Biomolecular Simulations: Methods and Protocols*: 529-578.

Jónsson, Hannes, Greg Mills, and Karsten W. Jacobsen. 1998. "Nudged elastic band method for finding minimum energy paths of transitions." In *Classical and quantum dynamics in condensed phase simulations*, pp. 385-404.

Mookherjee, Abhigya, Michael J. Van Stipdonk, and P. B. Armentrout. 2017. "Thermodynamics and reaction mechanisms of decomposition of the simplest protonated tripeptide, triglycine: a guided ion beam and computational study." *Journal of The American Society for Mass Spectrometry* 28 (4):739-757.

Lee, Jesi, Tobias Kind, Dean Joseph Tantillo, Lee-Ping Wang, and Oliver Fiehn. 2022. "Evaluating the accuracy of the QCEIMS approach for computational prediction of electron ionization mass spectra of purines and pyrimidines." *Metabolites* 12 (1):68.

Wang, Shunyang, Tobias Kind, Dean J. Tantillo, and Oliver Fiehn. 2020. "Predicting in silico electron ionization mass spectra using quantum chemistry." *Journal of cheminformatics* 12 (1):1-11.

Borges, Ricardo M., Sean M. Colby, Susanta Das, Arthur S. Edison, Oliver Fiehn, Tobias Kind, Jesi Lee et al. 2021. "Quantum chemistry calculations for metabolomics: Focus review." *Chemical reviews* 121 (10):5633-5670.

Wang, Shunyang, Tobias Kind, Parker Ladd Bremer, Dean J. Tantillo, and Oliver Fiehn. 2022. "Quantum Chemical Prediction of Electron Ionization Mass Spectra of Trimethylsilylated Metabolites." *Analytical Chemistry* 94 (3):1559-1

Pacific Northwest National Laboratory

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99354

1-888-375-PNNL (7665)

www.pnnl.gov