Projet de Recherche pour i-DREAM

Titre du projet/ Title of the project

Quantum-Inspired Neural Network Potentials and Simulations for Metabolomics

Acronyme/ Acronym

QIN-MET

CV des porteurs de chaque équipe (1 page max/personne)/CV of the PIs of each partner

Porteur de l'équipe du LCT : Dr. Riccardo Spezia, DR CNRS
Laboratoire de Chimie Théorique, Sorbonne Université, UMR 7616 CNRS
Email: riccardo.spezia@sorbonne-universite.fr
URL: https://www.lct.jussieu.fr/pagesperso/rspezia/

Professional Preparation

2000	Laurea (M.S.), Chemistry, Università di Roma "La Sapienza", Italy.
2004	Ph.D., Chemistry, Università di Roma "La Sapienza", Italy.
2012	Habilitation, Université d'Evry-Val-d'Essonne, France.

Academic/Professional Appointments

2018 - …	CNRS Research Director (DR2), Laboratoire de Chimie Théorique, Sorbonne Université, France.
2022-…	Director of the CECAM-FR-MOSER node, Maison de la Simulation, Saclay.
2006-2017	CNRS Researcher, LAMBE, Université d'Evry, France.
2005-2006	CNRS Postdoctoral Fellow, Université d'Evry, France.
2004-2005	CNRS Postdoctoral Fellow, Ecole Normale Supérieure Paris, France.

Publication record : about 150 papers in peer reviewed journals or book chapter, 24 invited talks at conferences, one book.

Synergistic Activities

2025-…	Member of the board of Chemistry PhD School, Università di Roma "La Sapienza", Italy.
2018-…	Head of the group Dynamics simulations: structure and reactivity at LCT.
2019-2024	General secretary of the Division de Chimie Physique (SCF, SFP).
2015-2019	Board of the Ile-de-France section of the Societé Chimique de France.
2015-2018	Direction committee of the Féderation de recherché Chimie Physique de Paris Saclay.

Awards and grants

2024-2025	InfrAnalytics project: *Ion-molecule complexes and synthesis of complex organic molecules via IR irradiation* at ICP, Université Paris-Saclay.

2024 Felix Beamline project: *Synthesis of organic molecules with astrochemical interest induced by IR irradiation of ion-molecule complexes*. Nijmegen (The Netherlands).
2022 Visiting professor, Università di Roma La Sapienza, Italy
2021-2025 ANR grant, consortium project (ENS, SU), responsible of the SU site.
2017 Visiting professor, Universidad de Valladolid, Spain.
2015 CNRS PEDR (*Prime d'encadrement doctorale et de recherche*) award.
2015-2018 ANR grant, ANR-NSF international program in chemistry
2010-2014 ANR young researcher grant. PI of the project.

5 significant publications relevant to the project

1. A.Martin-Somer, V.Macaluso, G.L.Barnes, L.Yang, S.Pratihar, K.Song, W.L.Hase and R.Spezia. *Role of Chemical Dynamics Simulations in Mass Spectrometry Studies of Collision-Induced Dissociation and Collisions of Biological Ions with Organic Surfaces.* J. Am. Soc. Mass Spectrom. 31, 2-24 (2020). https://doi.org/10.1021/jasms.9b00062

2. F.Angiolari, S.Huppert and R.Spezia. *Quantum versus Classical Unimolecular Fragmentation Rate Constants and Activation Energies at Finite Temperature from Direct Dynamics Simulations.* Phys. Chem. Chem. Phys. 24, 29357-29370 (2022). https://doi.org/10.1039/D2CP03809A

3. A.F.Perez-Mellor and R.Spezia. *Determination of Kinetic Properties in Unimolecular Dissociation of Complex Systems from Graph-Theory Based Analysis of an Ensemble of Reactive Trajectories.* J. Chem. Phys. 155, 124103 (2021). https://doi.org/10.1063/5.0058382

4. F.Angiolari, S.Huppert, F.Pietrucci and R.Spezia. *Environmental and Nuclear Quantum Effects on Double Proton Transfer in Guanine-Cytosine Base Pair.* J. Phys. Chem. Lett. 14, 5102–5108 (2023). https://doi.org/10.1021/acs.jpclett.3c00747

5. A.Carrà, V.Macaluso, P.W.Villalta, R.Spezia and S.Balbo. *Fragmentation Spectra Prediction and DNA Adducts Structural Determination.* J. Am. Soc. Mass Spectrom. 30, 2771-2784 (2019). https://doi.org/10.1007/s13361-019-02348-7

Porteur de l'équipe du IPCM : Dr. Denis Lesage, IR Sorbonne-Université

Professional Preparation
1990 DEA, Analytical Chemistry, INSTN, UPMC, France
1995 Ph.D., Analytical Chemistry, UPMC, France
2021 Habilitation, Sorbonne-Université, France

Academic/Professional Appointments
2022- … Research Engineer HS, Special level, Sorbonne-Université, France
2018-2022 Research Engineer HS, Sorbonne-Université, France
2007-2018 Research Engineer 1, UPMC, France
1995-2017 Research Engineer 2, UPMC, France
1995 (3 m) Engineer, CEA Saclay, France

Publication record: about 73 papers in peer reviewed journals or book chapter, 3 invited talks at seminars.

Synergistic Activities

2025-…	Group leader of the CSOB team (IPCM), Sorbonne-Université, France
2017-2024	Group co-leader of the CSOB team IPCM, Sorbonne-Université, France

Grants and management

2021	Organizing committee 38th Informal Meeting in Mass Spectrometry, Paris
2017	Scientific committee 35th Informal Meeting in Mass Spectrometry, Aussois
2012-…	Operational manager Infranalytics FR-2054 and TGE FT-ICR FR-3624
2008-…	Partners of 7 ANR, 2 Labex, 12 Soleil Beamline projects, 1 Clio project

5 significant publications related to the project

1) C.Chalet, D.Lesage, E.Darii, A.Perret, S.Alves, Y.Gimbert and J.C.Tabet. *Enantioselective Reduction of Noncovalent Complexes of Amino Acids with CuII via Resonant Collision-Induced Dissociation: Collision Energy, Activation Duration Effects, and RRKM Modeling.* J. Am. Soc. Mass Spectrom. 35, 456-465 (2024). https://doi.org/10.1021/jasms.3c00355

2) P.Bayat, D.Gatineau, D.Lesage, A.Martinez and R.B.Cole. *Benchmarking higher energy collision dissociation (HCD) by investigation of binding energies of gas-phase host–guest complexes of hemicryptophane cages.* J. Mass Spectrom. 57, e4879 1-9 (2022). https://doi.org/10.1002/jms.4879

3) D.Gatineau, H.Dossmann, H.Clavier, A.Memboeuf, L.Drahos, Y.Gimbert and D.Lesage. *Ligand effects in gold-carbonyl complexes: Evaluation of the bond dissociation energies using blackbody infrared radiative dissociation.* Int. J. Mass Spectrom. 463, 116545 (2021). https://doi.org/10.1016/j.ijms.2021.116545

4) P.Bayat, D.Lesage and R.B.Cole. *TUTORIAL: Ion activation in tandem mass spectrometry using ultra-high resolution instrumentation.* Mass Spectrom. Rev. 39, 680-702 (2020). https://doi.org/10.1002/mas.21623

5) D.Lesage, S.Mezzache, Y.Gimbert, H.Dossmann and J.C. Tabet. *Extended kinetic method and RRKM modeling to reinvestigate proline's proton affinity and approach the meaning of effective temperature.* Eur. J. Mass Spectrom. 25, 219-228 (2019). https://doi.org/10.1177/1469066718822054

Résumé du projet/Summary of the project

(5 lignes)

The project aims to extend present chemical dynamics simulations of collision-induced dissociation spectra by incorporating a machine learning approach based on neural network potentials trained on quantum calculations. The area of focus is metabolomics, which includes very large molecular diversity, and relies extensively on database search for identification. Quantum simulations will predict spectra and reaction mechanisms.

Support demandé/Project type

Thèse/PhD

Scientific description (5 pages max.)

Objectives and description of the project (4 pages environ)

The primary objective of this PhD project is to **develop a Machine Learning (ML)-based force field and apply it to the fragmentation dynamics of molecular ions**, in relation to tandem mass spectrometry (MS/MS) experiments. This approach is particularly relevant to metabolomics, which involves the identification and quantification of a vast number of metabolites lacking reference standards.

To simulate reactive dynamics, the MARS direct dynamics code (developed in the laboratory) will be modified and integrated with FENNOL, a Neural Network Potential (NNP) library also developed in-house. Quantum chemistry calculations will be used to train the NNP, allowing for reactive simulations that may also capture nuclear quantum effects. This approach will provide quantum-based reactive dynamics and predictive MS/MS spectra.

Using chemical dynamics trajectories to predict MS/MS spectra has proven to be a powerful tool over the past 15–20 years, initially developed by Hase and Spezia for collisions with inert gas atoms and later extended to thermal excitation by the same authors (as reviewed in a recent perspective article [1]). This approach was first applied using DFT or MP2 for very small systems [2] and later adapted to faster methods such as semi-empirical Hamiltonians (AM1, PM6, etc.) [3] and tight-binding DFT [4]. Other groups have since developed similar strategies; notably, Grimme and co-workers employed their tight-binding DFT approach to simulate fragmentation dynamics [5], while Wang, Fiehn, and co-workers recently proposed a similar method [6]. Each of these groups has developed its own code, but a major bottleneck remains the computational cost associated with the vast number of required trajectories (typically 1,000–10,000 for medium-sized molecules). In this project, we propose the development of a dedicated ML-based reactive NNP to overcome this limitation. This will enable us to expand our unique approach and distinguish from other research groups in Europe and the USA.

In the present project, simulations will focus on problem related to structural assignment relevant to metabolomics. Molecules present in a biological system under specific conditions, such as exposure to drugs, altered physiology and adaptation to environment are investigated by metabolomics. It relies on advanced analytical techniques for metabolite identification, with MS/MS spectra being one of the most widely used methods [7].

Nowadays, thousands of metabolites are identified by searching libraries or using *in silico* methods. However, spectral MS/MS libraries suffer from low reproducibility and high inter-instrument variability, despite significant scientific community efforts to improve identification protocols. Alternatively, *in silico* methods are developed for spectrum prediction [8,9]. Experimental MS/MS spectra are often acquired at high resolution to determine precise masses and infer reliable atomic compositions of precursor and fragment ions. However, the rapid advancement of experimental techniques has generated a vast amount of MS data, surpassing the progress in reliable compound identification methods and development of efficient software.

The chemical complexity of known metabolites is immense, and their number continues to grow as new environments and organisms are studied. An additional challenge in metabolite identification arises from their lower concentration in the sample [10]. Moreover, databases cover only a fraction of existing biomolecules, and the true extent of unknown metabolites remains uncertain. In non-targeted analyses, up to 95% of detected signals remain

unassigned when searching databases such as MassBank, NIST, METLIN, GNPS, PubChem, Human Metabolome Database (HMDB), etc. Compound identification can be improved by integrating complementary information, such as isotopic pattern analysis, chromatographic retention times, ion size measurements via ion mobility (or collision cross sections) [11], and NMR data. The most definitive method for compound characterization involves comparing all available experimental data with that of a synthetic reference molecule analysed under identical conditions. However, this approach is often prohibitively expensive or even unfeasible, **highlighting the need for advanced methodological, instrumental, and computational tools to improve metabolite identification**.

Our ultimate goal is to develop a computationally efficient method for simulating fragmentation dynamics to generate MS/MS spectra. Metabolomics presents an ideal application, as it focuses on small- to medium-sized organic molecules — well-suited targets for chemical dynamics simulations. These simulations will facilitate the understanding and prediction of spectra independently on experiments and, ultimately, enable **the creation of a fully *in silico* metabolomics database**. Unlike proteomics, where peptide sequence complexity is lower and fragmentations follow more predictable patterns, current *in silico* prediction methods for metabolomics, based on spectral comparisons or fragmentation trees, remain highly unreliable.

Furthermore, **the impact of the activation mode on the resulting spectra is highly under estimated**. The same system can yield different MS/MS spectra depending on the activation method used, which is strongly influenced by the mass spectrometer itself. This underscores **the need for collaboration between theoretical and experimental groups**. In simulations it is possible to model different activation modes by tuning initial conditions [1]. In fact, experimental MS/MS spectra depend not only on the type of mass spectrometer but also on the experimental conditions. The two primary factors that influence the content of an MS/MS spectrum are the internal energy transferred to the ion (and its evolution over time) and the decay time of the excited ion. To simplify modelling, activation-deactivation, and fragmentation processes are often treated as consecutive and independent events (Lindemann-Hinshelwood model). However, this simplification is not always applicable and requires more complex master equation modelling. For over 20 years, the CSOB team has been studying these activation and fragmentation processes using RRKM-QET theory, with mass spectrometry remaining under kinetic control due to the short decay times (typically ranging from μs to s) [12].

The CSOB team has extensive experience in measuring thermochemical quantities related to fragmentation, such as critical energy and transition state entropy. We employ several models to describe the evolution of internal energy during fragmentation [13]: i) the equilibrium thermodynamic temperature model for soft activation techniques like Blackbody Infrared Radiative Dissociation (BIRD) [14] and Low-Energy Collision-Induced Dissociation (L-E CID) [15]; ii) the equilibrium truncated thermal distribution model for small systems out of thermal equilibrium [16]; iii) the characteristic temperature model for high-energy activations like Higher Collision Dissociation (HCD) [17], accessible on the Orbitrap; and iv) the effective temperature model for the Threshold CID technique [18]. These four fragmentation techniques yield distinct MS/MS spectra, and each spectrum also depends on the collision energy in the laboratory reference frame. These factors are not accounted for when creating spectral databases, which explains the significant challenge of comparing an experimental spectrum with a reference spectrum obtained in another laboratory or by a different researcher. This also explains why multiple reference spectra for the same compound can exist in libraries. In summary, controlling time and energy is essential for obtaining reproducible and comparable spectra.

The comparison of experimental spectra with calculated spectra as well as with those produced by prediction algorithms such as SIRIUS [19] will be a key step in this work.

From a computational perspective, the novel aspect will be the development of a machine learning potential, which will be achieved through the FENNOL NNP approach [20]. To create the NNP specific to the fragmentation of molecular ions via direct dynamics, we will adopt a strategy that generates structures from fragmentation dynamics using previously employed methods, including semi-empirical Hamiltonians and tight-binding DFT. Analysis of the resulting structures will be conducted using graph theory for classification purposes. In this context, we will utilize a recently developed approach within our group [21]. The NNP will be constructed from these structures using higher-level quantum chemistry calculations (typically DFT) to derive the forces and other necessary properties. If necessary, the NNP will be refined iteratively via an active learning procedure. Specifically, for the fragmentation of ionic systems, the FENNOL approach will need to be extended to account for charged systems and to model interactions between charged species and dipolar species, as ion-molecule intermediates are common in gas-phase fragmentation. Since we will focus on the fragmentation of protonated molecules, which are closed-shell systems, quantum methods like DFT will be reliable in describing their reactivity. This NNP potential will then be validated and recalibrated by comparing the results with existing simulations and, ultimately, with experimental MS/MS spectra.

Once developed, the NNP will be **integrated into the MARS code** for simulations. Previously, theoretical MS/MS spectra were generated using the VENUS code, developed by the group of the late Prof. W.L. Hase (Texas Tech University, USA). The principal investigator (PI) of this project had a longstanding collaboration with this group and is currently one of the main supporters for users of the code. Unfortunately, this code cannot be modified and, furthermore, it is written in Fortran77, which makes it difficult to integrate with newer utilities. To address this limitation, the group has developed a new code over the past few years, MARS, written in modern Fortran90. This new code offers several advanced features and is fully developed at LCT. It also supports the Ring Polymer Molecular Dynamics (RPMD) method, enabling the inclusion of nuclear quantum effects. This will allow us to evaluate their role in fragmentation, especially in mechanisms involving one or more proton transfers [22].

We will apply this approach to study the MS/MS spectra of metabolites, gain insight into their fragmentation mechanisms, and ultimately enrich databases with new molecular information. Typical metabolites considered for this work are steroids, flavonoids, and nucleic acids. In previous works, simulations were performed on testosterone [23] and methyl-guanine [24], comparing the results with experiments conducted under specific activation conditions. These systems provide a practical starting point to develop a machine learning potential, as we can generate different structures along the reaction pathways and compare them with previous results. As demonstrated in the case of methyl-guanine fragmentation (see Figure 1), the theoretical spectrum obtained from chemical dynamics simulations is in good agreement with experimental results. In this case, simulations were performed using the relatively costly semi-empirical Hamiltonians. However, they could be conducted with the NNP potential, reducing computational cost while achieving the same (or better) quality.

This approach will be applied to systems studied experimentally by the IPCM group, such as flavonoids and estrogens. A key aspect will be integrating different activation modes into MARS to develop an efficient tool for theoretical mass spectrometry that is also accessible to experimentalists. The members of **LCT group** involved in the project are **Riccardo Spezia**, an expert in reaction dynamics, particularly in generating MS/MS spectra from chemical

dynamics simulations, and **Thomas Plé**, one of the main developers of the FENNOL approach.

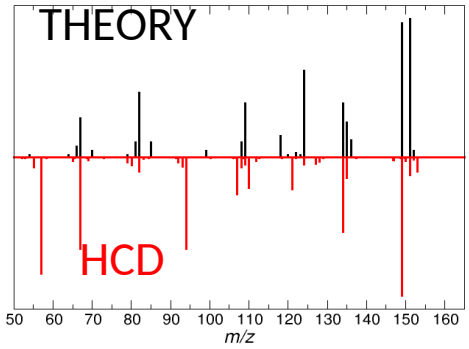

Figure 1. *Example of MS/MS spectra of methyl-guanine obtained from chemical dynamics simulations (in black) and HCD experiments (in red), extracted from Ref 24.*

Methyl-guanine and testosterone will be the first systems selected for developing NNP for fragmentation, given our prior experience with these molecules. Additionally, they will serve as a starting point for studying their respective molecular classes — nucleic acids and steroids. In particular, steroids will be a key focus of this project, as they provide an ideal case for effectively integrating simulations with experimental data. Our results can be used both to query existing databases and to build an *in silico* database that complements them.

On the strength of its experience in understanding and controlling fragmentation processes, **the CSOB team at IPCM** is heavily involved in the creation of spectral databases in the field of metabolomics (**Sandra Alves**, **Estelle Rathahao-Paris**, **Denis Lesage**). CSOB is developing a strong collaboration on this subject with the MNHN (A. Marie, A. Paris) in partnership with ChemBio France (CNRS). Above all, CSOB is also a partner in the National Metabolomics and Fluxomics Infrastructure (MetaboHub), which brings together the main players in the field of metabolomics (CNRS, CEA, INRAe, Inserm, INSA and four French universities).

One key factor in comparing calculations with experiments is the reaction time scale. In computational methods, this is typically limited to picoseconds, although it can be extended to nanoseconds using NNP potentials. However, in experiments, the ion decomposition time is significantly longer. In HCD, the most widely used MS/MS technique for metabolite detection, it ranges from microseconds to milliseconds, far exceeding the timescales accessible in calculations. As a result, in computed spectra, simple cleavage mechanisms tend to dominate over rearrangement fragmentation processes, which are strongly influenced by entropic effects.

To address these discrepancies, RRKM modelling of energy-resolved mass spectra (ERMS) will be employed to better interpret the differences between experimental and theoretical spectra. This modelling will be conducted for both HCD and L-E CID, allowing excitation times to be adjusted to favour the study of rearrangement fragmentation processes.

In this project, we aim to bridge the gap between simulations and experiments through three complementary approaches:

1. **Developing NNP potentials** to enable significantly longer simulation times.
2. **Modulating ion decomposition times in experiments** as a function of fragment abundances.

3. **Modulating activation energy in simulations** as a function of fragment abundances.
4. **Establishing a close interplay between calculations and experiments**: when an isomer of the initial structure or a primary fragment identified in simulations is considered a potential precursor for a key fragmentation (from experimental results and/or RRKM modelling), it will be re-used as the initial structure for a new set of fragmentation dynamics.

Notably, this iterative procedure is an extension of the approach used to account for the formation of different tautomers in chemical dynamics simulations of collisional activation in MS/MS, which can lead to distinct fragmentation patterns [24,25].

Feasibility. Risk assessment and management (1/2 page environ)

The most challenging and innovative aspect of this project is the developing of a Machine Learning-based potential to model the reactivity of molecular ions in the gas phase. To achieve this, we will generate a large dataset through direct dynamics simulations based on standard methods. If successful, this approach would represent a major breakthrough in the field. In this process, we can face many problems and the development can be more complex than expected. However, even if integrating FENNOL proves complex, assessing its feasibility would still be a significant outcome, paving the way for potential refinements or methodological adjustments. Also, in the worst case, i.e. FENNOL cannot deal with fragmentation, we can either move to other ML schemes (like e.g. the Deep Potential-Smooth Edition (DeepPot-SE) model [26], which is implemented in the DeePMD-kit package. [27]) and in any case understand which are the key physical or informatics factor at the basis of the problem.

The reliability of NNP will be constantly check through comparison with more standard theoretical calculations and experiments.

Position of the project in regard to the iDream objectives (1/2 page environ)

This project aims to establish innovative methods for using quantum chemistry to predict fragmentation spectra. Developing Neural Network Potentials for reactive systems is at the forefront of current research in quantum chemistry, as it significantly reduces the computational cost of propagating trajectories while maintaining high accuracy, thanks to training on high-level calculations. Additionally, this approach will enable the investigation of the potential role of nuclear quantum effects in fragmentation dynamics.

Furthermore, its application to metabolomics will provide a novel tool to assist scientists working on therapeutic challenges where metabolite analysis and identification are crucial.

The project is built on a collaboration between the theoretical group at LCT, specialized in reaction dynamics, and the experimental group at IPCM, expert in mass spectrometry. Thanks to this synergy, facilitated by i-DREAM, theory and experiments will be closely integrated, enhancing the impact of results from both perspectives. Moreover, the project will extend collaboration to research groups at MNHN and CEA, specialized in metabolomics, strengthening interactions that could support future funding opportunities through ANR or other national and international initiatives.

Références / Bibliography

1) A.Martin-Somer, V.Macaluso, G.L.Barnes, L.Yang, S.Pratihar, K.Song, W.L.Hase and R.Spezia. *Role of Chemical Dynamics Simulations in Mass Spectrometry Studies of Collision-Induced Dissociation and Collisions of Biological Ions with Organic Surfaces.* J. Am. Soc. Mass Spectrom. 31, 2-24 (2020).

2) R.Spezia, J.-Y.Salpin, M.-P.Gaigeot, W.L.Hase and K.Song. *Protonated Urea Collision-Induced Dissociation. Comparison of Experiments and Chemical Dynamics Simulations.* J. Phys. Chem. A 113, 13853-13862 (2009).

3) E.Rossich Molina, D.Ortiz, J.-Y.Salpin and R.Spezia. *Elucidating collision induced dissociation products and reaction mechanisms of protonated uracil by coupling chemical dynamics simulations with tandem mass spectrometry experiments.* J. Mass Spectrom. 50, 1340–1351 (2015).

4) R. Nieman, R. Spezia, B. Jayee, T. Minton, W. L. Hase and H. Guo. *Exploring Reactivity and Product Formation in N(4S) Collisions with Pristine and Defected Graphene with Direct Dynamics Simulations.* J. Chem. Phys. 153, 184702 (2020).

5) J.Koopman and S.Grimme. *From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics.* J. Am. Soc. Mass Spectrom. 32 (7), 1735-1751 (2021).

6) J.Lee, D.J.Tantillo, L.-P.Wang, and O.Fiehn. *Predicting Collision-Induced-Dissociation Tandem Mass Spectra (CID-MS/MS) Using Ab Initio Molecular Dynamics.* J. Chem. Inf. Model. 64, 7470-7487 (2024).

7) S.Alseekh et al., *Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices.* Nature Methods, 18, 747-756 (2021).

8) H.Oberacher et al. *A European proposal for quality control and quality assurance of tandem mass spectra Libraries.* Sci Eur. 32(1) 43 (2020).

9) N.Paulhe et al. *PeakForest: a multi-platform digital infrastructure for interoperable metabolite spectral data and metadata management.* Metabolomics 18:40 (2022).

10) S.L.Collins, I.Koo, J.M.Peters, P.B.Smith, and A.D.Patterson. *Current Challenges and Recent Developments in Mass Spectrometry–Based Metabolomics.* Annu. Rev. Anal. Chem. 14, 467-487 (2021).

11) A.Delvaux, E.Rathahao-Paris, S.Alves. *Different ion mobility-mass spectrometry coupling techniques to promote metabolomics.* Mass Spectrometry Reviews 41(5) 695-721 (2021).

12) R.A.Marcus, O.K.Ric. *The Kinetics of the Recombination of Methyl Radicals and Iodine Atoms.* J. Phys. Chem. 55, 894−908 (1951).

13) D.Lesage, S.Mezzache, Y.Gimbert, H.Dossmann and J.C.Tabet. *Extended kinetic method and RRKM modeling to reinvestigate proline's proton affinity and approach the meaning of effective temperature.* Eur. J. Mass Spectrom. 25, 219-228 (2019).

14) R.C.Dunbar. *BIRD (blackbody infrared radiative dissociation): Evolution, principles, and applications.* Mass Spectrom. Rev., 23(2), 127-158 (2004).

15) C.Chalet, D.Lesage, E.Darii, A.Perret, S.Alves, Y.Gimbert and J.C.Tabet. *Enantioselective Reduction of Noncovalent Complexes of Amino Acids with CuII via Resonant Collision-Induced Dissociation: Collision Energy, Activation Duration Effects, and RRKM Modeling.* J. Am. Soc. Mass Spectrom. 35, 456-465 (2024).

16) D.Gatineau, H.Dossmann, H.Clavier, A.Memboeuf, L.Drahos, Y.Gimbert and D.Lesage. *Ligand effects in gold-carbonyl complexes: Evaluation of the bond dissociation energies using blackbody infrared radiative dissociation.* Int. J. Mass Spectrom. 463, 116545 (2021).

17) P.Bayat, D.Gatineau, D.Lesage, A.Martinez and R.B.Cole. *Benchmarking higher energy collision dissociation (HCD) by investigation of binding energies of gas-phase host–guest complexes of hemicryptophane cages.* J. Mass Spectrom. 57, e4879 1-9 (2022).
18) P.B.Armentrout. *Mass spectrometry—not just a structural tool: the use of guided ion beam tandem mass spectrometry to determine thermochemistry.* J. Am. Soc. Mass Spectrom. 13, 419-434 (2002).
19) K.Dührkop at al. *SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information.* Nature Methods 16 299–302 (2019).
20) T.Plé, O.Adjoua, L.Lagardère and J.-P.Piquemal. FeNNol: *An efficient and flexible library for building force-field-enhanced neural network potentials.* J. Chem. Phys. 161, 042502 (2024).
21) A.F.Perez-Mellor and R.Spezia. *Graph Determination of Kinetic Properties in Unimolecular Dissociation of Complex Systems from Graph-Theory Based Analysis of an Ensemble of Reactive Trajectories.* J. Chem. Phys. 155, 124103 (2021).
22) V.H.Wysocki, G.Tsaprailis, L.L.Smith, L.A.Breci. Mobile and localized protons: a framework for understanding peptide dissociation. J. Mass Spectrom. 35, 1399-1406 (2000).
23) G.Lee, E.Park, H.Chung, Y.Jeanvoine, K.Song, and R.Spezia. *Gas phase fragmentation mechanisms of protonated testosterone as revealed by chemical dynamics simulations.* Int. J. Mass Spectrom. 407, 40-50 (2016).
24) A.Carrà, V.Macaluso, P.W.Villalta, R.Spezia and S.Balbo. *Fragmentation Spectra Prediction and DNA Adducts Structural Determination.* J. Am. Soc. Mass Spectrom. 30, 2771-2784 (2019).
25) A. Carrà and R. Spezia. *In silico tandem mass spectrometer: an analytical and fundamental tool.* Chemistry-Methods 1, 123-130 (2021).
26) L.Zhang, J.Han, H.Wang, W.Saidi, R.Car and W.E. *End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems.* Adv. Neural Inf. Process. Syst. 31 (2018).
27) H.Wang, L.Zhang, J.Han and W.E. *DeePMD-kit: A Deep Learning Package for Many-Body Potential Energy Representation and Molecular Dynamics.* Comput. Phys. Commun. 228, 178-184 (2018).

Work program (1 page max.)

Research program (1 page max.)

Our working program is here detailed and divided in four main steps.

Step 1: Creation of simulated MS/MS and experimental database and identification of the synergies between them. We will then select some key systems which will be used to train the machine learning (ML) potential. We will also run some standard simulations using semi-empirical quantum chemistry calculations (semi-empirical Hamiltonians, like PM6, and/or tight-binding DFT). Here we will do also some experiments on prototypical systems of interest in metabolomics, such as steroids and/or flavonoids.

Step 2: Development of the ML potential based on the FENNOL approach. This is probably the most difficult and delicate step. We will need to enlarge the present data-set to charged (protonated) systems, running quantum chemistry calculations (at DFT level) and comparing results with both experiments and simulations. Trajectories will also be used to generate a large data-set of structures to learn the Neural Network used in FENNOL. The dataset will then be iteratively extended using an active learning procedure. Finally, the comparison with simulation results will be crucial since we can have exactly the same activation conditions and see how the comparison varies as a function of, for example, activation energy.

Step 3: Once the ML potential is completed and tested, we can couple the ML engine with the simulation activation conditions we have in our chemical dynamics codes, presently VENUS and MARS. Specifically, we will implement it in MARS, since it is written in modern Fortran90 and it is developed at LCT. Activation conditions contain: (I) thermal activation; (ii) collisional activation. In the last case, we can use the analytical interaction potentials developed in the past by the LCT group. The inclusion of the ML engine in our MARS code will easily allow it. In this step, we will tune our activation modes with the spectrometers present at the IPCM group. Three activation modes, often giving different MS/MS fragmentation spectra, will be used: TCID, HCD and L-E CID. These involve collisional activation on target gases of different sizes, leading to different internal energy distributions deposited on the precursor ions. Fragmentation spectra will be also simulated using RRKM modelling to extract thermochemical fragmentation data (entropy and critical energy related to the transition state).

Step 4: A stronger correlation between the databases will be undertaken. From one side, we will explore in details the cases in which there is a discrepancy between the different experimental set-ups and the spectra present in the database. We will be able to the run fragmentation simulations using the developed ML potential to better investigate the database and eventually adding new spectra coming from simulations also as a function of the activation conditions. Among the numerous libraries of fragmentation spectra, we will mainly use HMDB (the Human Metabolome Database), MONA (MassBank or North America) and Pubchem. In collaboration with MNHN, we also have access to hundreds of perfectly characterized metabolites. It will also be possible to compare calculated spectra with that generated by software such as SIRIUS and as well as spectra found with molecular network approach.