



UNIVERSITY OF TRENTO

**Bridging Experiments and Simulations to
Address Computational Challenges in Atom
Probe Tomography and Structural Biology**

THESIS AUTHOR & PHD CANDIDATE:
GIOVANNI NOVI INVERARDI

SUPERVISORS AND ADVISORS:
PROF. GIANLUCA LATTANZI, PHD. SIMONE TAIOLI

Statistical and Biological Physics Laboratory,
Department of Physics, University of Trento

Submitted on the 15th July 2026 for the degree of Doctor of Philosophy

Abstract

This thesis investigates the fundamental mechanisms governing the application of an experimental technique, Atom Probe Tomography (APT) to the field of structural biology through a combination of atomistic simulations and theoretical modeling. In fact, APT typically is utilized to determine the properties of metals and alloys, but recently has got a huge leap forward towards new fields of application, among these, the protein structure determination field. Thus, the primary objective of this manuscript is to improve the reliability and interpretability of APT analyses by flanking the experimental findings with theoretical calculations, providing a predictive framework to rationalize the APT workflow at the nanoscale.

A first part of the work focuses on the interaction between silica and biomolecular systems. Using classical molecular dynamics simulations, structural and dynamical properties of these hybrid interfaces are characterized, highlighting how adsorption phenomena influence the local protein behavior, with consequences on the dynamics of biomolecular features. These results provide insight into the challenges of analyzing organic–inorganic systems with APT.

The study then focuses on the evaporation mechanisms of cations from non-conductive nano-samples under the influence of ultrafast terahertz pulses. By employing *ab initio* molecular dynamics simulations, this work highlights the role of transient electric/laser fields in triggering field evaporation, offering a possible framework to better understand experimental observations and to extend APT applicability to insulating materials.

A further contribution of this thesis lies in the investigation of defects in sol–gel silica and their impact on APT analysis. Through computational analysis concerning Density Functional Theory calculations, it is proposed that structural impurities significantly improve the conductivity of the APT silica specimen, underlining the importance of accounting for material imperfections when trying to develop a framework to analyze insulators like silica within the APT workflow.

Overall, this thesis aims at bridging theoretical calculations and experimental findings, elucidating the key role that *in silico* methods might play as complementary techniques to provide predictive insights into the main procedures, phases and processes of APT. The findings presented here help to advance the understanding of APT in challenging new frontiers and open novel perspectives for its application to structural biology.

Acknowledgements

First, I would like to thank my supervisor Gianluca Lattanzi for his guidance, support, and trust throughout the course of my PhD. His availability for discussion was essential in shaping the direction and quality of this research. Beyond his academic supervision, his encouragement and patience played a fundamental role in helping me navigate the challenges encountered along the way.

I would like to acknowledge the MIMOSA project for its essential support throughout the development of this work. The project provided the financial and institutional framework that made this research possible, ensuring access to the necessary computational resources, collaborative networks, and scientific infrastructure. Beyond its practical support, MIMOSA fostered a highly stimulating and interdisciplinary environment, bringing together researchers with complementary expertise and encouraging continuous exchange between theoretical and experimental approaches. This context played a fundamental role in shaping the direction and quality of the research presented in this thesis.

I would also like to express my sincere gratitude to the MIMOSA project team and its wider community. Their constant support, openness, and collaborative spirit created an exceptionally stimulating environment in which ideas could be freely exchanged and developed. In particular, I am grateful for the valuable discussions, feedback, and encouragement provided throughout the project, which significantly contributed to both the scientific progress and my personal growth as a researcher. Being part of this group has been a truly enriching experience. For this reason, my thanks extend all the PhD students and postdocs of MIMOSA, particularly to Matteo De Tullio, Francesco Carnovale, Gustav Eriksson, Tommaso Morresi.

Generally, I also want to express my sincere gratitude to all my colleagues: Giovanni Marini, Costanza Paternoster, Alessio Bartocci, Lorenzo Petrolli, Andrea Corradini, Stefano Mocatti and many others. Their constant availability for discussion, willingness to share ideas, and constructive feedback were invaluable throughout the development of this work. The collaborative atmosphere they fostered made it possible to tackle complex problems with greater clarity and confidence, and contributed significantly to both the scientific outcomes and my personal growth as a researcher.

Preface

The research presented in this thesis was supported by the MIMOSA project (<https://cordis.europa.eu/project/id/101046651>), which provided the essential financial resources for its development. Beyond enabling access to computational facilities and experimental collaborations, this funding also supported my position and allowed me to dedicate my efforts to the research activities described in this work. In this sense, the MIMOSA project was not only a scientific framework, but also the foundation that promoted a highly stimulating environment, where young and senior scientists had the opportunity to exchange ideas, engage in constructive discussions, and develop collaborative approaches to complex research problems. This dynamic setting fostered both intellectual growth and interdisciplinary interaction, significantly enriching the quality and perspective of the work presented in this thesis.

The research presented here reflects a progressive consolidation of themes, methods, and questions that have been explored across multiple studies (theoretical and experimental), each contributing to shaping the perspective adopted in this manuscript. This thesis emerges from a research path developed at the intersection of computational modeling and structural biology in a specific context of an experimental technique: Atom Probe Tomography. In fact, I had the opportunity to collaborate with several researchers with an experimental background: the group of A. Vella (University of Rouen - France) and the group of Martin Andersson (Chalmers University of Technology, Gothenberg - Sweden). The main goal of MIMOSA is to determine individual structures of proteins (i.e. with single-molecule resolution) utilizing the Atom Probe Tomography framework. This represents the first attempt to apply such technique to the protein structure determination field. During my PhD, the MIMOSA project fostered an interdisciplinary environment where theoretical modeling and experimental validation could inform each other in a continuous feedback loop.

This thesis, therefore, should be seen not as a collection of isolated studies, but as a coherent effort to address a broader question: how to improve the reliability and resolution of atom probe tomography when applied to structural biology.

Beyond its scientific contributions, this work represents a significant stage in my academic development. It reflects the challenges of navigating complex systems, the necessity of adapting methodologies to evolving questions, and the value of collaboration in tackling interdisciplinary problems. While the conclusions drawn here try to promote specific answers, they also open new directions for future research, reinforcing the idea that understanding at the atomic scale remains an evolving and collective endeavor.

The thesis is structured in six main Chapters. Chapter 1 contains a large but detailed introduction to the protein structure determination field, Atom probe Tomography, the MIMOSA project and also the role of computational modeling within this wide and broad context. Chapter 2 contains a general description of the main computational methods utilized in this manuscript, explaining the main features and parameters needed to understand the results of my PhD work regarding MIMOSA, stored in Chapters 3, 4 and 5. Chapter 3 is based on the these three papers:

Giovanni Novi Inverardi, Francesco Carnovale, Lorenzo Petrolli, Simone Taioli, and Gianluca Lattanzi. **Silica In Silico: A Molecular Dynamics Characterization of the Early Stages of Protein Embedding for Atom Probe Tomography.** *Biophysica*, 3(2):276–287, 2023.

Giovanni Novi Inverardi, Lorenzo Petrolli, Francesco Carnovale, Alessio Bartocci, Simone Taioli, and Gianluca Lattanzi. **Adsorption of silica oligomers on biomolecules: Structural and dynamical insights for atom probe tomography via classic molecular dynamics simulations.** *Computational and Structural Biotechnology Journal*, 27:2537-2543, 2025.

Gustav Eriksson, David Mayweg, Giovanni Novi Inverardi, Francesco Carnovale, Mats Hulander, Mattias Thuvander, Lorenzo Petrolli, Gianluca Lattanzi, Simone Taioli, Martin Andersson (2026). **Silica-Embedded Proteins Reveal Molecular Scale Structure by Deep-UV Laser Atom Probe Tomography.** *In submission & peer reviewing phases.*

Chapters 4 and 5 are based on the following journal papers, respectively:

Matteo De Tullio, Giovanni Novi Inverardi, Michella Karam, Jonathan Houard, Marc Ropitiaux, Ivan Blum, Francesco Carnovale, Gianluca Lattanzi, Simone Taioli, Gustav Eriksson, Mats Hulander, Martin Andersson, Angela Vella, and Tommaso Morresi. **Evaporation of cations from nonconductive nanosamples using single-cycle terahertz pulses: an experimental and theoretical study.** *Physical Review B*, 111(4):045428, 2025.

Gustav Eriksson, Matteo De Tullio, Francesco Carnovale, Giovanni Novi Inverardi, Tommaso Morresi, Jonathan Houard, Marc Ropitiaux, Ivan Blum, Emmanuel Cadel, Gianluca Lattanzi, Mattias Thuvander, Martin Andersson, Mats Hulander, Simone Taioli, and Angela Vella. **Role of Defects in Atom Probe Analysis of Sol-Gel Silica.** *ACS Omega*, 10(30):33741-33754, 2025..

Chapters 3, 4 and 5 share a similar structure: there is always a first introductory section, followed by two sections describing the aim of the study and the simulation protocol. Then, a large section storing all the produced data is present, followed by a last conclusive chapter, sometimes also offering some perspectives of the work. Lastly, Chapter 6 sums up every result presented, providing the lecturer with a general view of the whole thesis. Moreover, the thesis contains also Appendix Sections: Specifically, segments A and B store all the Supplementary Data and Information of Chapters 3 and 4, while the last unit 7 describes also other side projects on which I worked outside of the MIMOSA.

Contents

List of figures	xiv
1 Introduction	1
1.1 Protein Structure Determination Field: State of the Art	1
1.2 The Relevance of Local Chemistry: from Cryo-EM to Atom Probe Tomography .	4
1.3 Overview of Atom Probe Tomography	5
1.3.1 Key Concepts to understand Atom Probe Tomography	6
1.3.2 Principles of the Ion-Field Evaporation	8
1.3.3 Modulation of the Detection Rate in Atom Probe Tomography	11
1.4 Laser-assisted Atom Probe Tomography	12
1.4.1 Photo-induced ionization	13
1.4.2 Thermal Field Evaporation	15
1.5 Atom Probe Tomography applied on biological materials	17
1.5.1 The Relevance of the Sample Preparation Protocol in Atom Probe Tomography applied on Biology	17
1.5.2 The First Application of Atom Probe Tomography on Proteins	18
1.5.3 Details of the Sample Preparation Protocol and the FIB-SEM Procedure	23
1.5.4 Details of the Laser-Assisted APT Analysis	26
1.6 The MIMOSA Project	29
1.7 The Role of <i>in Silico</i> Modeling in the MIMOSA Project	31
1.7.1 Molecular Mechanics Methods Might Unravel Key Mechanisms of the Sample Preparation Protocol	32
1.7.2 First Principles Calculations Might Clarify Crucial Aspects of the Ion-Field Evaporation in APT analyses	32
2 Foundations of Molecular Mechanics and First-Principles Simulations	35
2.1 Introduction to Molecular Dynamics	35
2.1.1 Assumptions of Molecular Dynamics Simulations	35
2.1.2 The Born-Oppenheimer Approximation	36
2.2 The Molecular Dynamics Algorithm	37
2.2.1 Classical Force Fields	38

2.2.2	Reactive Force Fields	45
2.2.3	Integrators	47
2.2.4	Periodic Boundary Conditions and Simulation Box	50
2.2.5	Thermostats and Barostats	50
2.3	Coarse-Grained Molecular Dynamics	53
2.3.1	Motivation for Coarse-Grained Models	54
2.3.2	Principles of Coarse-Graining	54
2.3.3	Dynamics in Coarse-Grained Models	55
2.3.4	Parameterization Strategies in Coarse-Grained Models	55
2.3.5	The MARTINI Force Field	57
2.4	Introduction to Density Functional Theory	61
2.4.1	From the Exact Many-Body Problem to the Kohn–Sham Formalism	61
2.4.2	The Exchange–Correlation Functional and the Local Density Approximation (LDA)	64
2.4.3	Algorithm of the DFT Procedure	66
2.5	Introduction to Real Time, Time Dependent Density Functional Theory	67
2.5.1	Foundations of Time-Dependent Density Functional Theory: the Time-Dependent Kohn–Sham Equations	67
2.5.2	Real-Time Propagation in Time Dependent Density Functional Theory	69
2.5.3	Coupled Electron–Nuclear Dynamics: TDDFT and the Ehrenfest Approximation	71
2.5.4	Algorithm for First-Principles Molecular Dynamics with RT-TDDFT and Ehrenfest	73
2.6	Density Functional Theory and Linear Optical Response	74
2.6.1	Introduction to Optical Absorption and Linear Response Theory	74
2.6.2	Density Functional Theory, Linear Optical Response and Absorption Spectra	78
2.6.3	From DFT and Linear Response Theory to the Computation of Absorption Spectra	80
3	Atomistic Insights into Silica–Biomolecule Interactions for APT sample preparation via MD simulations	83
3.1	Motivating MD Simulations: Recent Experimental Studies on GFP <i>via</i> APT	83
3.2	The MD Study	86
3.2.1	Aim of The Study	86
3.2.2	Biomolecular Systems	86
3.3	Simulation Protocol	88
3.3.1	Crystal Structures	88
3.3.2	Force Fields Selection and Compatibility	89
3.3.3	MD setup	91
3.3.4	Details about Analysis and Parameters	92
3.4	Results of MD Simulations	94
3.4.1	The Waterglass System: Assessment of the Silica Behavior in Water	94
3.4.2	The Monomeric and Mixed Scenarios: Adsorption Patterns, Structural Determinants, and Domain-Specific Selectivity	100

3.4.3	Effects of the Silica Concentration on the Protein Dynamics	113
3.5	Concluding Remarks	117
3.6	Perspectives of the Study: Going Beyond Classical Molecular Dynamics	118
3.6.1	Modeling Reactivity: Reactive Molecular Dynamics and Reaxff	119
3.6.2	Coarse Grained Molecular Dynamics and Reactivity: A New Hope	122
4	Unraveling THz-mediated Ion Field Evaporation in APT via <i>ab initio</i> Molecular Dynamics	133
4.1	Introduction to THz-assisted Atom Probe Tomography	133
4.1.1	THz Radiation Properties	133
4.1.2	THz-assisted Field Evaporation and Thermal Effects	134
4.1.3	THz-assisted APT and Field Enhancement	135
4.1.4	The Single-Cycle THz Pulse Polarity in APT	137
4.2	Motivating <i>ab initio</i> MD Simulations: Anomalous THz Pulse Polarity Behavior in THz-assisted APT of Silica Tips	138
4.3	The <i>ab initio</i> MD Study	143
4.4	Simulation Protocol	144
4.4.1	Software Details	144
4.4.2	External Field	144
4.4.3	Time Propagation Scheme	148
4.4.4	Si(OH) ₄ Atomic Coordinates and Simulation Box Parameters	148
4.4.5	Optimization under Static Electric Fields	149
4.4.6	Boundary Conditions	149
4.4.7	Production Run Details	150
4.5	Results of <i>ab initio</i> MD Simulations	151
4.5.1	Modeling the THz-assisted Field Evaporation Mechanism in Molecules: The First Model	151
4.5.2	Bridging the Gaps between Bulks and Molecules: The Second Model	157
4.5.3	The True Role of Electron Dynamics in the Context of Negative Pulse Polarity and APT	163
4.5.4	Near-Apex Field Enhancement in the Context of Negative Pulse Polarity and APT	167
4.6	Concluding Remarks	168
5	Understanding The Role of Defects During APT Analyses of Sol-gel Silica through DFT Calculations	171
5.1	Introduction to the Modeling of the Optical Properties of Defects in SiO ₂ Matrices	171
5.1.1	The Role of Defects within SiO ₂ Matrices	171
5.1.2	Optical Linear Response and Defects Modeled <i>via</i> Density Functional Theory	173
5.2	Motivating DFT Calculations: Experimental Data on Defects of SiO ₂ Matrices within the APT Context	176
5.2.1	SiO ₂ Matrices in APT Display Defects	176

5.2.2	Defects Alter the Absorption Properties of the SiO ₂ Matrices utilized in APT	178
5.3	The DFT Study	180
5.4	Simulation Protocol	181
5.4.1	DFT and Optical Properties	181
5.4.2	Exchange–Correlation Functional Utilized	182
5.4.3	External Electric Field	184
5.4.4	Generation and Optimization of Amorphous Structures and Defective Structures	184
5.5	Results of Optical Absorption Evaluations through DFT	187
5.5.1	Preliminary Analyses on the Amorphous SiO ₂ Models	187
5.5.2	Defects Play a Major Role in Optical Response Compared to the APT Static Field	189
5.5.3	Optical Response of Typical Defects Present in APT Analyses	191
5.6	Conclusive Remarks	197
6	Conclusions	199
7	Other Projects	203
	Appendices	207
A	Supplementary Data of Chapter 3	207
A.1	Supplementary Figures	207
A.2	Simulation Protocol for the Reproduction of the Dupuis <i>et al.</i> Results	209
A.2.1	Software and Initial Configuration	209
A.2.2	Force Field	209
A.2.3	Energy Minimization and Equilibration	209
A.2.4	Production Simulation	209
A.2.5	Hybrid GCMC–PT Scheme for Reactive MD	209
A.3	Modeling the pH in Molecular Dynamics	212
A.3.1	Fundamental Concept of Constant-pH Simulations	212
A.3.2	Discrete Constant-pH Molecular Dynamics	212
A.3.3	Continuous Constant-pH Molecular Dynamics	213
A.3.4	Coupling Constant-pH Methods with Replica Exchange	215
A.3.5	Applicability of CpHMD to the Systems in this Thesis	215
B	Supplementary Data of Chapter 4	217
B.1	Supplementary Figures	217
C	Case of Study: The YTHDF1-Ebselen Compound Interaction	221
C.1	Introduction	221
C.1.1	Covalent Drugs	221
C.1.2	The Ebselen Compound	222
C.1.3	Ebselen as a covalent inhibitor of YTHDF proteins	223

C.1.4	The Relevance of the $\beta 4$ – $\beta 5$ Loop of YTHDF1 for the Inhibitory Mechanism	224
C.2	Aim of The Study	227
C.3	Simulation Protocol	228
C.3.1	Crystal Structures	228
C.3.2	Force Fields Selection and Compatibility	228
C.3.3	Plain MD setup	229
C.3.4	Localized Volume-Based Metadynamics	230
C.4	Preliminary Results and Conclusions	233
References		235

List of Figures

1.1	A Scheme of X-Ray Crystallography	2
1.2	A Scheme of Cryo-Electron Microscopy	3
1.3	A Scheme of Atom Probe Tomography	7
1.4	Reconstruction Process of Atom Probe Tomography	8
1.5	The Field Evaporation Mechanism in APT	9
1.6	The Photo-Ionization Mechanism in APT	14
1.7	The Thermal Evaporation Mechanism in APT	16
1.8	The Gel-like Silica Matrix	19
1.9	IgG Aggregates During the Sol–Gel Process	20
1.10	TEM Micrograph and 3D Reconstruction of an IgG Antibody	21
1.11	Atom Probe Mass Spectra of an IgG Molecule Embedded in Si(OH) ₄ -derived Matrix	21
1.12	Workflow of the First Atom Probe Tomography Analysis on Proteins	22
1.13	The Sample Preparation Protocol	23
1.14	The FIB-SEM Procedure	25
1.15	LPE Analysis	28
2.1	Vibrational Energy Levels of a General Diatomic Molecule	39
2.2	Schematic Representation of Bonded Interactions	40
2.3	Fourier Decomposition of the Torsional Energy for Rotation about the C–C Bond of N-Butane	42
2.4	Typical Box Shapes Utilized in Molecular Dynamics Simulations	50
3.1	APT Applied on the Green Fluorescent Protein	84
3.2	Organic and Inorganic Molecules Studied in this Work	87
3.3	Starting Configuration of the Monomeric and Mixed Scenarios.	89
3.4	Hierarchical Clustering Workflow	93
3.5	DFT Calculations on the Si(OH) ₄ Molecular Structure	95
3.6	Radial Distribution Functions (Waterglass System)	96
3.7	Jensen–Shannon Divergence Heatmap	98
3.8	Evaluation of the Solvent Accessible Surface Area	101
3.9	Radial Distribution Functions of Monomeric and Mixed Systems	102

3.10	Covering Factor of EGFP in the Monomeric and Mixed Systems	103
3.11	3D Isodensity Maps of Silica Species	104
3.12	Hydrogen Bonds between the Proteins and Silica Species	105
3.13	Heatmaps of Protein–Silica Contacts	107
3.14	Root Mean Square Fluctuations of Protein–Silica Contacts	108
3.15	Hierarchical clustering analysis of the $A\beta_{1-40}$ dynamics	110
3.16	Hierarchical clustering analysis of the IL22RA1 dynamics	112
3.17	Results of the Alternative Mixed Scenario	114
3.18	EGFP at Diverse Concentrations of $\text{Si}_3\text{O}_{10}\text{H}_8$	115
3.19	The ReaxFF Force Field	119
3.20	Reproduction of a Polymerizing Silica Matrix <i>via</i> ReaxFF	121
3.21	Reproduction of a Polymerizing Silica Matrix <i>via</i> the CG Sticky MARTINI Frame- work	124
3.22	Classical Coarse-Grained MD Simulations Replicating the Mixed System for ubiq- uitin	125
3.23	The three CG Models for a Si-O-Si Bond within the MARTINI 3 Framework . .	128
3.24	Results of Model 3 within the reactive MARTINI 3 Framework	131
4.1	The Terahertz Laser Frequency	134
4.2	The Antenna Effect: THz versus UV	136
4.3	The THz Pulse Polarity	137
4.4	THz Pulse Polarity Results from APT on SiO_2 and LaB_6 samples	140
4.5	Scheme of the Orientation of each Field within the APT Chamber	142
4.6	The First Model	144
4.7	Model of a Single-Cycle THz Pulse	145
4.8	Values of THz Laser Amplitudes and DC Field	146
4.9	The Static Electric Field	147
4.10	Optimization of the $\text{Si}(\text{OH})_4$ Geometry under a Static Electric Field	149
4.11	Results of the First Model	153
4.12	Visualization of the Evaporation Dynamics in a $\text{Si}(\text{OH})_4$ Molecule	155
4.13	The Second Model	158
4.14	Results from The Second Model (1)	160
4.15	Results from The Second Model (2)	161
4.16	Results from The Second Model (3)	163
4.17	Electron-assisted Atom Probe Tomography	165
5.1	Comparison UV/deep-UV Lasers in APT Sol-Gel Silica Analyses	177
5.2	Comparison of Mass Spectra of UV/deep-UV Lasers in APT Sol-Gel Silica Analyses	178
5.3	Absorbance of SiO_2 Matrices Utilized in APT	179
5.4	The three SiO_2 Matrices Utilized for the Evaluation of Optical Properties <i>DFT</i> .	181
5.5	Performances of Two Exchange-Correlation Functionals	183
5.6	One Amorphous SiO_2 Matrix Utilized for the Evaluation of Optical Properties <i>DFT</i>	185

5.7	One Amorphous SiO ₂ Matrix Utilized for the Evaluation of Optical Properties <i>DFT</i>	188
5.8	Complex Dielectric Function and Refractive Index of the Three Amorphous SiO ₂ Models	189
5.9	Amorphous SiO ₂ with and without Oxygen Defects Models	190
5.10	Amorphous SiO ₂ Carbon Defects Models	191
5.11	Amorphous SiO ₂ Oxygen and Hydrogen Defects Models	193
5.12	Amorphous SiO ₂ Oxygen and Sodium Defects Models	194
5.13	Amorphous SiO ₂ Oxygen and Gallium Defects Models	195
A.1	RMSD of IL22RA1	207
A.2	Dendrogram of the Hierarchical Clustering Analysis on IL22RA1	208
A.3	Covering Factor of Ubiquitin and SUMO-1 in the Monomeric and Mixed Systems	208
B.1	APT Setup for the SiO ₂ and LaB ₆ Samples	217
B.2	Generation of the THz Pulse	218
B.3	Results from The Second Model (4)	219
C.1	The ebselen Compound Interacting with a Cysteine in the SARS-CoV-2 main protease (M ^{Pro})	223
C.2	Crystal structure of the YTH domain of the YTHDF1 protein bound to m6A . .	224
C.3	Ebselen covalently binds YTHDF1	225
C.4	Comparison of the $\beta 4$ – $\beta 5$ loop conformations in YTHDF1.	226
C.5	Comparison of the three PDB files representing three different states of the $\beta 4$ – $\beta 5$ loop	227
C.6	The reconstructed $\beta 4$ – $\beta 5$ loop in 7PCU	229

1.1 Protein Structure Determination Field: State of the Art

The comprehension of the mechanisms that underlie protein functions in living cells is a central milestone of molecular biology. The most detailed information is obtained from three-dimensional (3D) structures, which directly reflect the chemical characteristics of proteins and ultimately govern their biological activity. The access to 3D structural information is therefore essential for interpreting fundamental biochemical phenomena, including enzymatic catalysis and the interactions between drugs and their targets. Techniques capable of producing quantitative spatial representations at the atomic resolution must then draw on a wide array of experimental approaches, collectively addressing structure, composition, mechanics, and dynamics across many length scales.

The last century saw a strong development of such powerful tools. Historically, X-ray crystallography has been the first methodology to produce an efficient 3D characterization of macromolecules at near atomic resolution through the analysis of diffraction patterns from highly ordered crystals [1]. Its main workflow is illustrated in Fig. 1.1. Briefly, the sample is induced to precipitate, and the formed crystals are exposed to an X-ray beam. The resulting diffraction patterns are then processed by molecular modeling algorithms to retrieve the so-called 3D electron density maps, ultimately needed to obtain spatial information on the protein of interest. X-ray crystallography requires a reliable source of highly pure and homogeneous crystals, frequently produced under non-physiological conditions. Such conditions depend on many variables: type of precipitant and its concentration, buffer and its pH, protein concentration, temperature, and possible inclusion of additives. An improper treatment of these often destabilizes sensitive samples, favoring the emergence of non-native conformations, or neglecting relevant flexible regions of the molecule. Moreover, intermolecular contacts within the crystal lattice can distort the backbone and side-chain geometries and restrict native dynamics, thus complicating functional interpretation [2]. As a result, crystallization remains the main bottleneck, particularly for membrane proteins and large conformational clusters and complexes [3, 4]. However, X-ray crystallography continues to provide highly accurate structural coordinates that

underpin mechanistic studies and structure-based ligand design.

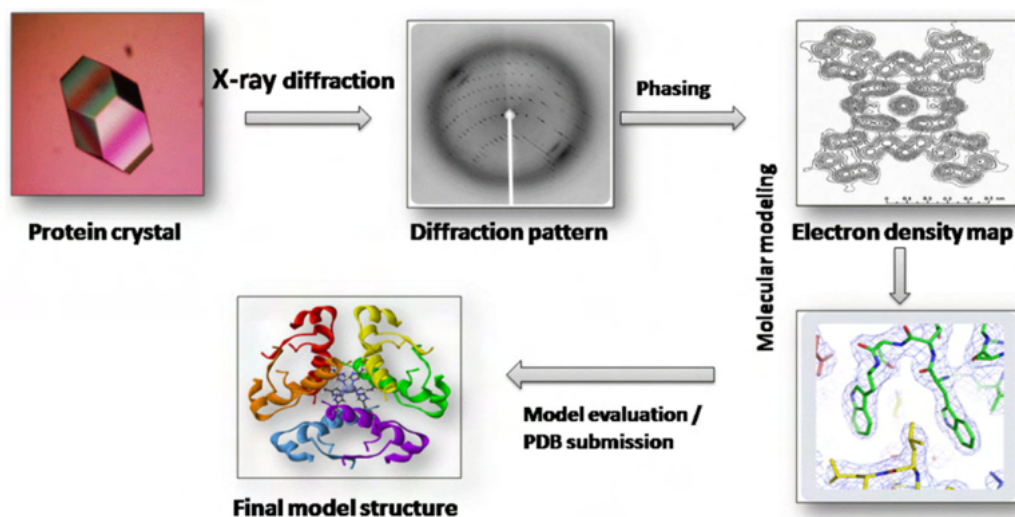


Figure 1.1: Schematic and typical workflow of the X-ray crystallography technique. Reproduced from Ref. [5]

Other procedures have been developed to complement data coming from X-rays. For example, Nuclear Magnetic Resonance (NMR) offers details on the local chemical environment of atoms within a protein that could be considered as constraints during reconstruction phases. In solution, NMR can uncover both structural and dynamic features with high specificity through the use of isotopic labeling and multidimensional experiments. Protonation states, ion binding, and information about rotamers are features that remain largely inaccessible to most other experimental techniques. Furthermore, NMR enables the study of proteins at physiological and ambient temperatures, allowing detailed, site-specific investigation without imposing harsh conditions that might alter their natural dynamics. On the other hand, NMR detection operates on objects with small molecular weight (40–50 kDa) and typically requires large amounts of relatively pure sample in the order of several milligrams [6, 7].

The true revolution in the field of structural biology can be represented by the advent of cryogenic electron microscopy (cryo-EM), developed between the 80s-90s as method to study proteins in a quasi-native state at sub-nanoscale resolution [8]. The placement of a cryo-EM sample inside a transmission electron microscope (TEM) is shown in Fig. 1.2. A small volume of purified protein solution is applied onto a metal grid, typically copper, often coated with a perforated amorphous carbon film. Excess liquid is removed by blotting with a paper fiber, and the grid is then rapidly plunged into liquid ethane to flash-freeze the sample. This procedure ideally

produces a thin layer of vitreous, non-crystalline ice in which proteins are distributed in multiple orientations. Achieving this vitreous state is critical, as a slower freezing at higher temperatures would cause sample expansion and structural damage. During imaging in the transmission electron microscope (TEM), the frozen grid is maintained at liquid nitrogen temperature. The resulting images, obtained through the holes in the carbon film, capture two-dimensional (2D) projections of individual protein complexes in diverse orientations. These numerous 2D projections are subsequently combined to generate a single 3D reconstruction, revealing the complete 3D architecture of the protein [9, 10].

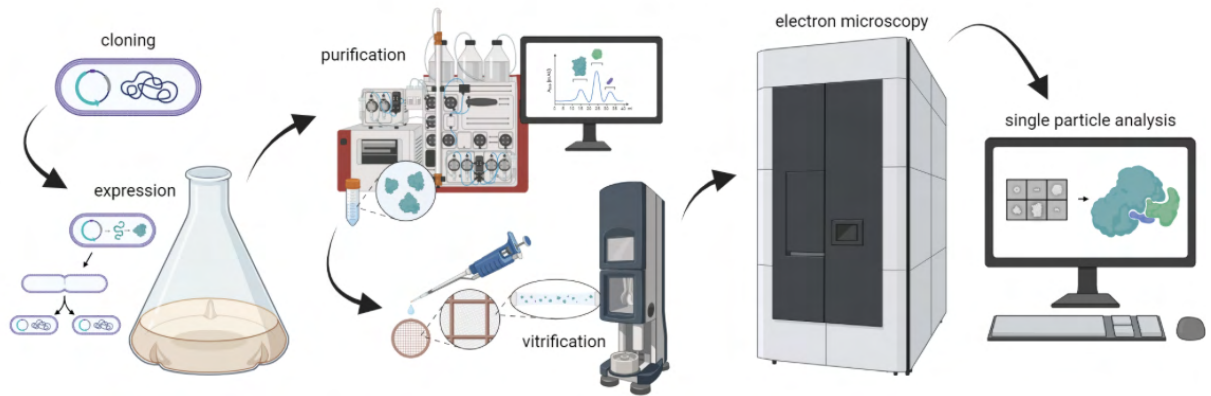


Figure 1.2: Schematic diagram displaying the workflow of the Cryogenic Electron Microscopy. Created in Biorender [11]

Driven by major advances in electron detection and image analysis, cryo-EM has emerged as a technique that rivals X-ray crystallography in achievable resolution. By 2008, near-atomic structures of several viruses and ribosomes had already been determined, with resolutions on the order of 6 Å [12, 13]. These early achievements were enabled by decades of progress in microscope hardware, sample preparation, and computational methods. A decisive turning point occurred in 2013 with the commercial availability of direct electron detectors, marking the beginning of a rapid transformation of the field. This new generation of detectors delivers images of exceptional quality, while modern image-processing algorithms compensate for beam-induced motion and enable classification of particles into distinct conformational states. Together, these developments produce density maps with sufficient detail to infer atomic models for a broad range of biological specimens.

The enhanced image quality provided by direct detectors allows particle orientations to be assigned with greater accuracy and substantially improves the separation of heterogeneous structural states. In parallel, the extension of maximum-likelihood approaches within an empirical Bayesian framework has simplified image processing by enabling key parameters to be inferred

directly from the data, reducing the need for manual tuning. Consequently, over the past few years, numerous research groups have reported cryo-EM reconstructions at atomic or near-atomic resolution across diverse systems. Complexes with molecular masses below 200 kDa have become accessible targets, and achievable resolutions now routinely approach 3 Å, with values below 2 Å attainable under favorable conditions [14, 15]. Despite these advances, high-resolution structure determination by cryo-EM remains challenging due to the inherently low contrast of biological specimens, which are primarily composed of light elements (carbon, hydrogen, oxygen, nitrogen, phosphorus, and sulfur). This limitation leads to reduced signal-to-noise ratios, complicates 3D reconstruction, and necessitates careful management of radiation damage during electron-beam exposure. In addition, particle orientation and incomplete angular sampling can bias reconstructions, motivating approaches such as stage tilting or the use of engineered support films to increase orientational diversity [16].

The TEM is an instrument that can be also coupled with electron energy-loss spectroscopy (EELS) and energy-dispersive X-ray spectroscopy (EDS). EELS is particularly sensitive to low- Z elements, enabling detection of carbon, nitrogen, and oxygen edges, as well as analysis of near-edge fine structure associated with bonding, which in principle allows discrimination between chemical states [17, 18]. In contrast, EDS performs best for medium to heavy elements and offers rapid, wide-field elemental mapping. The electron beam that underpins TEM’s high spatial resolution also imposes a fundamental limitation for soft and biological materials due to radiation damage. Even under cryogenic conditions, the permissible electron dose is finite, enforcing trade-offs between signal-to-noise ratio, spatial resolution, and tilt-series completeness; gains in one dimension inevitably come at the expense of sample integrity [19]. In summary, TEM provides exceptional structural resolution and, when combined with EELS and EDS, enables chemically sensitive imaging. However, under realistic dose limits, achieving quantitative 3D stoichiometry of light elements at nanometric resolution still remains a significant challenge.

1.2 The Relevance of Local Chemistry: from Cryo-EM to Atom Probe Tomography

Although cryo-EM now routinely provides near-atomic 3D structures of biomolecules, understanding biological function *in situ* requires complementary chemical information. In particular, structural data must be integrated with spatially resolved, quantitative measurements of local composition at comparable length scales. While existing techniques excel at structural characterization, methods capable of delivering nanoscale, 3D maps of light-element chemistry (C, N, O) in biological specimens remain limited. In other words, the next generation techniques should gather data storing information about 3D structures *plus* local chemistry as an orthogo-

nal dimension of analysis. Various techniques have been tested during the recent years, in order to find promising candidates capable of adding an effective contribute to cryo-EM studies on biological specimens.

Atom Probe Tomography (APT) holds the potential to complement and extend the insights obtained from cryo-EM [20, 21]. The core principle of APT relies on the field-evaporation of individual ions from a needle-shaped specimen (radius less than 100 nm) under a strong electric field (typically greater than $1 \text{ V} \cdot \text{\AA}^{-1}$) combined with short, repetitive laser pulses. A position-sensitive detector provides lateral (X–Y) resolution of each field-evaporated ion, while the order in which ions hit the detector defines the depth (Z) coordinate. The time-of-flight information on each evaporated atom reveals its mass-to-charge-state ratio, i.e. its chemical nature. Consequently, APT enables the reconstruction of 3D compositional maps with sub-nanometer spatial resolution and absolute counting statistics [20, 22, 23]. Under favorable conditions, APT achieves parts-per-million sensitivity and can resolve isotopes, especially for light elements (C, N, O). Hydrogen detection is also possible, though accurate quantification requires careful management of background signals, adsorbates, and experimental artifacts [24].

Historically, APT emerged from field ion microscopy (FIM) introduced by Mueller in 1951. The atom probe field ion microscope (APFIM, 1967) combined FIM geometry with time-of-flight mass spectrometry, enabling chemical identification of field-evaporated ions [20]. A major milestone was the development of the local electrode atom probe (LEAP, 1996), which established the modern APT instrument architecture [25]. The introduction of laser pulsing expanded the applicability of APT beyond good conductors, allowing routine analysis of semiconductors, oxides, and minerals [25, 26]. More recently, ultrafast terahertz (THz) pulses have been explored to assist APT analyses in conductive materials, although the limits of this approach, particularly for low-conductivity specimens, remain under investigation. APT and EM can be integrated effectively in advanced materials workflows. Cryo-EM or TEM can first provide high-resolution structural maps and identify regions of interest. These regions can then be extracted via focused ion beam lift-out to create site-specific specimens for APT. In such combined workflows, EM delivers detailed structural imaging under dose-optimized conditions but is limited in quantifying light elements, whereas APT provides precise, 3D chemical information, including atomic-level stoichiometry, at the nanoscale.

1.3 Overview of Atom Probe Tomography

This section introduces some key concepts about the Atom Probe Tomography workflow, also discussing the principles of the field-evaporation mechanism.

1.3.1 Key Concepts to understand Atom Probe Tomography

The core concept of Atom Probe Tomography relies on field-induced evaporation: atoms at the apex of a needle-shaped specimen become ionized and are expelled (or evaporated) under a strong electric field generated by a steady DC voltage combined with short triggering pulses (either high-voltage or laser). The DC bias produces a concentrated field at the nanoscale apex, while the added pulse momentarily boosts the local field to the level needed for ions to be emitted. A scheme of the procedure is reported in Fig. 1.4. The shape of the tip is non-casual, as the specimen must be sufficiently sharp to concentrate the electrostatic field at its apex. In a spherical approximation, the surface potential and field are $V = \frac{Q}{4\pi\epsilon_0 R}$ and $F = \frac{Q}{4\pi\epsilon_0 R^2}$, which immediately gives $F = V/R$. In practice, real emitters are more accurately modeled as a hemispherical cap atop a conical shank within an electrode assembly. The influence of this geometry and the surrounding boundary conditions is captured by a specimen-instrument field factor k_F (commonly denoted K_f), leading to the practical working relation:

$$F = \frac{V}{k_F R}, \quad (1.1)$$

with k_F typically 2-8 and dependent on shank angle, apex shape, and extractor coupling [27, 28, 29, 30, 31]. The electric fields required for field ionization or evaporation are on the order of 10^{10} V m^{-1} (equivalent to tens of V nm^{-1}), which can be readily achieved for tip radii of a few to tens of nanometers under kilovolt-scale biases. Due to the rapid decay of the field away from the apex, small variations in the tip geometry or in the local electrostatic environment can significantly influence ion trajectories and the accuracy of tomographic reconstruction.

The evaporation phase begins as follows: a brief pulse (typically high-voltage pulses for good conductors) transiently raises the electrostatic field at the tip apex above the binding energy of surface atoms, causing them to be ejected and accelerated toward the detector. The mass-to-charge ratio (MCR) of each ion is determined from its time-of-flight (ToF) using energy conservation [32]:

$$\text{MCR} = \frac{m}{n} = \frac{2eV \text{ToF}^2}{L^2}, \quad (1.2)$$

where m is the ion mass, n its charge state, $V = V_{\text{DC}} + V_{\text{Pulse}}$ the total accelerating potential (generally the potential difference between tip apex and detector), e the elementary charge, and L the flight path length. Accurate APT analysis requires careful specimen preparation: the sample is sharpened into a tip with radius $R \lesssim 100 \text{ nm}$ to achieve the necessary fields [33]. Standard preparation methods include electro-polishing and focused ion beam (FIB) milling [34]. Modern position-sensitive, time-resolving detectors record both the impact coordinates and the time-of-flight [35], enabling high-fidelity 3D reconstructions of atomic composition and

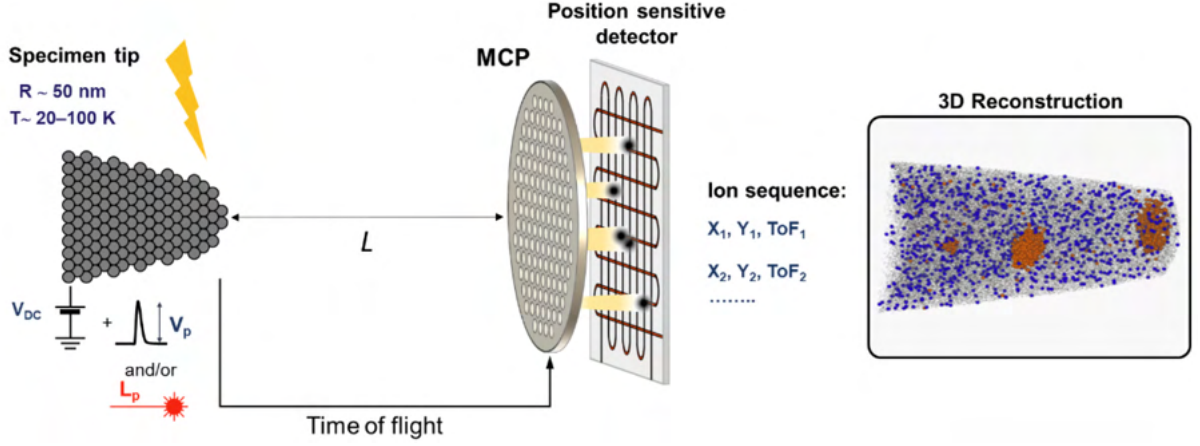


Figure 1.3: Schematic diagram displaying the Atom Probe Tomography experimental setup. The sample is sharpened to an ultra-fine tip (radius 100 nm), which is then subjected to a high electric field under illumination of a pulsed laser beam to induce field evaporation of the surface ions. A position-sensitive detector reconstructs the 3D profile of the original by coupling positional information and chemical nature of each evaporated ion.

structural organization. The reconstruction phase is a quite complex procedure: ion trajectories are typically modeled as rays originating from a projection point P located along the tip axis at a distance $OP = \xi R$ from the apex, where R is the local radius of curvature and ξ is the image-compression factor [36]. For a detector positioned at a distance L with $L \gg \xi R$, the point-projection magnification is given by:

$$M_{\text{proj}} = \frac{D}{d} = \frac{L + \xi R}{\xi R} \approx \frac{L}{\xi R}, \quad (1.3)$$

allowing the lateral specimen coordinates to be recovered by demagnifying the detector coordinates:

$$x = \frac{X_D}{M_{\text{proj}}}, \quad y = \frac{Y_D}{M_{\text{proj}}}. \quad (1.4)$$

Because atoms are initially placed on the tangent plane at the apex, a spherical curvature correction projects them back onto the cap of radius R :

$$dz' = R \left(1 - \sqrt{1 - \frac{x^2 + y^2}{R^2}} \right) \approx \frac{x^2 + y^2}{2R}, \quad (1.5)$$

which is valid for small emission polar angles ($r = \sqrt{x^2 + y^2} \ll R$). In practice, the depth of

the i -th ion is then accumulated as:

$$z_i = \sum_{j=1}^i dz + dz'_i, \quad (1.6)$$

where dz represents the mean depth increment derived from the sequence of evaporated ions. This procedure ensures an accurate 3D reconstruction of the specimen geometry and composition.

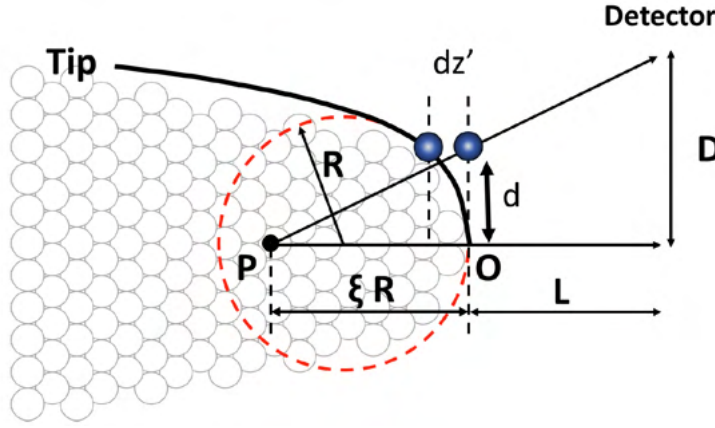


Figure 1.4: Point-projection geometry (adapted from [22]). The projection point P is located on the tip axis at a distance $OP = \xi R$ from the apex, where R is the local radius of curvature. The detector is positioned at a distance L from the apex, with $L \gg \xi R$.

1.3.2 Principles of the Ion-Field Evaporation

Field-evaporation is a complex mechanism, that entails the removal of an atom from a solid surface under the action of a strong external electric field (and also a laser pulse in some occasions), as schematically reported in Fig. 1.5 (a). An applied electric field polarizes surface atoms and progressively lowers the bonding barrier that confines them to the lattice. When the field becomes sufficiently strong, an atom can detach from the surface, transfer one or more electrons back to the solid, and accelerate away as a positively charged ion.

Early theoretical descriptions treated field evaporation as a thermally activated process in which atoms surmount a field-lowered energy barrier [37, 38, 39]. Gomer's charge-draining model emphasizes a gradual loss of electronic charge as the atom moves away from the surface [38]. In contrast, in the Müller's image-hump model, the atom is assumed to remain neutral until it reaches a critical distance from the surface, where it suddenly becomes fully ionized. Despite

their conceptual differences, both models allow evaporation to proceed via thermal activation, with quantum tunnelling becoming relevant when the barrier is sufficiently thin.

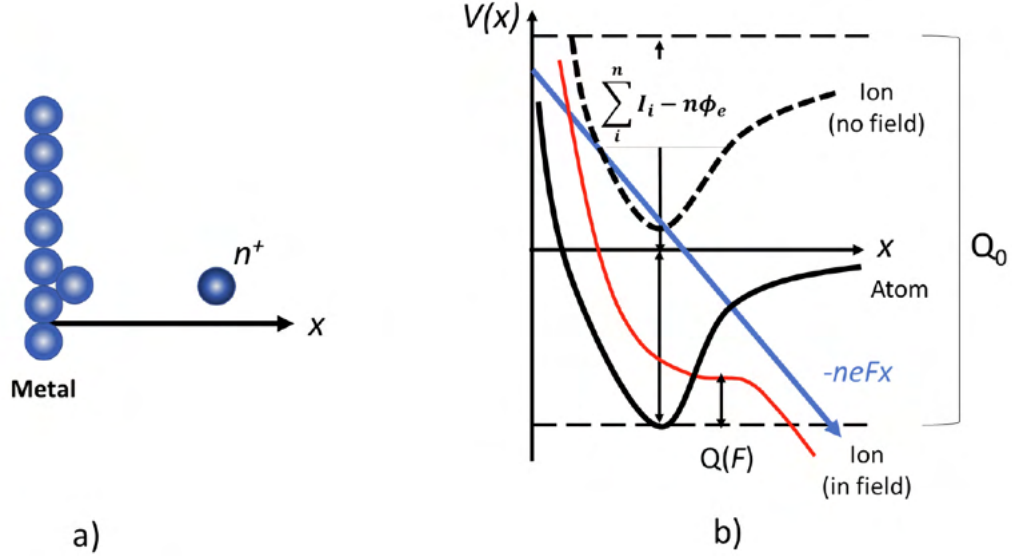


Figure 1.5: The field evaporation mechanism (adapted from [22]). (a) Field evaporation process; a cation detaches from a metal surface. (b) Atomic and ionic potential-energy diagrams with and without an applied electric field. V is the potential that each ion senses at a distance x from the surface without (black dashed line) and with (red line) the external electric field. The term $-neFx$ represents the work performed by the electric field on an ion of charge state n . I_i is the i -th ionisation energy, and ϕ_e the work function. Q_0 and $Q(F)$ are the energy barriers without and with field, respectively.

Here, the image-hump expression refers to the energy barrier an ion must overcome to leave the surface, created by the competition between the external electric field pulling it out of the material and its image charge pulling it back. Therefore, the external electric field lowers the ionic potential energy, while image-charge forces attract the ion back toward the surface. As a result, a field-dependent Schottky barrier generates (often referred to as the image hump), which the atom must overcome to evaporate [39].

Specifically, the associated potential-energy landscape sensed by an ion with (red line) and without (black dashed) the field is shown in Fig 1.5 (b). The term $-neFx$ represents the work performed by the electric field on an ion of charge state n as it moves away from the surface (the negative sign indicates that the electric field lowers the potential energy with increasing distance, promoting ion emission). The zero-field evaporation energy Q_0 represents the cost of removing an atom from the surface and ionizing it in the absence of an external electric field,

and is given by:

$$Q_0 = \Lambda + \sum_{i=1}^n I_i - n\phi_e, \quad (1.7)$$

where Λ is the binding (or sublimation) energy of the neutral atom at the surface, I_i are the successive ionization energies required to form an ion of charge state n , and ϕ_e is the electronic work function of the emitter. The last term accounts for the energy recovered when the n electrons released during ionization are returned to the metal.

In presence of the external field F , the field-dependent evaporation energy barrier $Q(F)$ for an ion of charge state n is reduced by the work performed by the field and by the image-charge interaction, yielding:

$$Q(F) = Q_0 - neFx - \frac{n^2e^2}{16\pi\epsilon_0x}, \quad (1.8)$$

where Q_0 is the zero-field evaporation energy, with Λ the surface binding energy of the neutral atom, I_i the successive ionization energies to produce a charge n ion, and ϕ_e the work function of the metal; x is the distance of the ion from the surface; e is the elementary charge and ϵ_0 the vacuum permittivity. The second term represents the energy reduction due to the applied electric field, while the third term accounts for the attractive interaction with the image charge in the metal. This construction captures the field-lowered activation barrier that must be overcome for field evaporation to occur.

Using the definition of $Q(F)$ above, the maximum field-lowered barrier can be obtained by evaluating the distance x at which $Q(F)$ is minimized. By differentiating Eq. (1.8) with respect to x and setting $dQ/dx = 0$, the critical distance where the ion is most likely to escape the surface is $x_{\text{crit}} = \sqrt{\frac{ne}{16\pi\epsilon_0F}}$. Substituting this optimal x_{crit} back into Eq. (1.8) yields a compact expression for the field-dependent evaporation energy in terms of the applied field F :

$$Q(F) = Q_0 - \sqrt{\frac{n^3e^3F}{4\pi\epsilon_0}}. \quad (1.9)$$

that explicitly shows how the combined effects of the external electric field and the image-charge interaction reduce the effective barrier, providing a simple yet quantitative link between the zero-field evaporation energy Q_0 and the field-assisted ion emission rate.

The initial models cited before (refs. [37, 38]) assumed that atoms could be emitted directly as multiply charged ions, but as a matter of fact, they were unable to reproduce the experimentally observed distributions of ion charge states. Subsequent experimental and theoretical work demonstrated instead that evaporation predominantly produces singly charged, or at most doubly charged, ions. In fact, higher charge states are generally formed after emission via post-ionization processes occurring in the strong near-surface field. Precisely, Haydock and Kingham

proposed the concept of post-ionization [40, 41]: atoms are initially emitted predominantly as singly charged ions (occasionally doubly charged) and may undergo further ionization in the strong local electric field, potentially through successive steps. Analogous to field ionization, the additional ionization from charge state n to $n + 1$ occurs at a critical distance x'_c from the surface:

$$x'_c = \frac{I_{n+1} - \phi_e}{F}, \quad (1.10)$$

where I_{n+1} is the $(n + 1)$ -th ionization energy, ϕ_e is the work function, and F is the local electric field. The probability of post-ionization is therefore both field-dependent and influenced by the ion's residence time in the ionization zone, which depends on its velocity and mass. Consequently, isotope-dependent charge-state distributions have been observed [42, 43]. Quantitative agreement with the measured field-dependent charge states has been demonstrated for rhodium [40, 41], and Kellogg showed that for tungsten, reproducing the average charge state requires a multi-step ionization sequence, inconsistent with single-step or simple charge-hopping models [44]. Kingham later tabulated the charge-state probabilities $P_n(F)$ for numerous elements, known as the Kingham curves [41], which form the basis of standard practice in field-evaporation analysis [39, 45].

To conclude, recent quantum-mechanical treatments based on first-principles and density-functional theory have been applied to field evaporation [46, 47, 48, 49, 50, 51, 52, 53]. While these approaches provide valuable microscopic insight into electronic structure and charge transfer, their quantitative predictive power remains an active topic of discussion, particularly regarding their applicability to realistic tip geometries and high-field conditions [54].

1.3.3 Modulation of the Detection Rate in Atom Probe Tomography

The evaporation rate at the tip of the specimen depends strongly on both the local electric field and the temperature. In the Arrhenius formalism, it is expressed as

$$\Phi_{\text{evap}} = \nu_0 N_{\text{king}} \exp \left(- \frac{Q(F)}{k_B T} \right), \quad (1.11)$$

where ν_0 is the attempt frequency, N_{king} is the number of high-field sites, $Q(F)$ is the field-lowered activation barrier, and T is the emitter temperature. An increase in the local field F (which reduces Q) or in T results in an exponential rise in the evaporation rate.

Arrhenius behavior has been widely observed [44, 55, 42], with typical attempt frequencies $\nu_0 \sim 10^{11}$ – 10^{13} Hz, sometimes weakly dependent on F , and measured activation energies below 1 eV [55]. At very low temperatures and for light ions, quantum tunneling contributes, producing a weak temperature dependence [39, 43, 42].

The critical evaporation field F_{EV} can be defined as the field at which the barrier would vanish, and given the expression for $Q(F)$ in 1.9:

$$Q(F_{\text{EV}}) = 0 \quad \Rightarrow \quad F_{\text{EV}} = \frac{4\pi\epsilon_0 Q_0}{n^3 e^3}, \quad (1.12)$$

For fields below the critical value, it is often convenient to use a linearized approximation of the barrier, commonly employed in Arrhenius-type descriptions of the evaporation rate:

$$Q_{\text{norm}}(F) \approx Q_0 \left(1 - \frac{F}{F_{\text{EV}}} \right). \quad (1.13)$$

This form directly links the exact inverse-field dependence of $Q(F)$ to a normalized, field-scaled barrier $Q_{\text{norm}}(F)$, facilitating its use in rate equations and experimental analysis. Combining eqs. 1.11 and 1.13 yields a compact relation between the field needed to sustain a given evaporation rate and the specimen temperature:

$$\frac{F}{F_{\text{evap}}} \approx 1 + \ln \left(\frac{\Phi_{\text{evap}}}{\nu_0} \right) \frac{k_B T}{Q_0}, \quad (1.14)$$

that displays a linear-in- T trend, observed across different materials [44, 55]. Thus, the field required to maintain a chosen detection rate is extremely sensitive to the surface temperature due to the $K_B T$ factor in eq. 1.14. To ensure ions are synchronized with pulsed excitation, inter-pulse evaporation must remain negligible (requiring a sufficiently low DC field at base temperature) in order to avoid information loss, while the pulsed excursion, either high voltage (HV) or laser-induced thermal, must be sufficient to evaporate relevant species predominantly during the pulse. In HV pulsing, this is controlled via the pulse fraction; in laser pulsing, via the temperature excursion at nearly constant DC field. If the DC field (or base T) is too high, low-field species spontaneously evaporate between pulses, causing composition bias and reduced depth fidelity. If the pulsed excursion is too small, high-field species remain, leading to sequence distortions [22, 33, 44].

Consequently, careful calibration of field and temperature, validated against element-specific curves, is essential for accurate atom-probe measurements.

1.4 Laser-assisted Atom Probe Tomography

As reported in previous sections, in APT atoms at the apex of the tip become ionized and are evaporated under a strong electric field generated by a steady DC voltage combined with short triggering pulses, either high-voltage or laser. This section describes the foundations of the latter

case, when APT is coupled with lasers, i. e. laser-assisted APT (la-APT).

Since the late 1970s, short laser pulses have been employed to trigger field evaporation in atom probe instruments. Early work by Tsong and co-workers established the role of photon-assisted field ionization processes [56, 57, 58]. Shortly thereafter, Tsong and Kellogg implemented the first pulsed-laser source on an atom probe microscope, demonstrating the feasibility of laser-triggered evaporation [59]. Numerous instrumental and methodological variants followed [60, 61], culminating in modern systems that employ sub-nanosecond [62], picosecond [63], and femtosecond laser sources [64]. Compared with high-voltage pulsing, laser pulsing significantly extends APT to poorly conducting materials, including semiconductors, oxides, and polymers. In metallic systems, it can also reduce the energy spread of emitted ions and improve mass resolving power, while simultaneously lowering electrostatic and mechanical stresses and reducing the likelihood of specimen fracture. Although several physical mechanisms may contribute to laser-assisted evaporation—including photo-ionization, transient field enhancement, and thermally assisted processes—it is now widely accepted that, for pulse durations $\gtrsim 100$ fs, la-APT is predominantly governed by thermal effects.

Following laser absorption, the electronic subsystem is heated almost instantaneously, while the lattice remains initially out of equilibrium. Over the electron–phonon coupling time, typically ranging from several hundred femtoseconds to a few picoseconds, energy is transferred to the lattice, with characteristic timescales that depend on both material properties and tip geometry. For laser pulses longer than this coupling time (typically on the order of 1 ps), the electronic and lattice subsystems are effectively coupled during the pulse, leading to a rapid rise in surface temperature. Under the presence of a strong DC electric field, this transient temperature increase promotes thermally assisted field evaporation [56, 57].

In conclusion, the field evaporation in la-APT reflects the interplay of optical, thermal, and electronic effects. In the following two sections, two relevant and key contributors in the ion-field evaporation process will be elucidated: *(i)* resonant photo-ionization of surface atoms and *(ii)* thermally assisted evaporation via a transient temperature rise.

1.4.1 Photo-induced ionization

Photo-ionization is a resonant process in which one or more photons excite an electron from a surface adsorbed atom (adatom) to an unoccupied state, thereby ionizing the adatom [65, 66]. Tsong and colleagues proposed three distinct mechanisms [22]. *(i)* Direct tunneling: in this process, an electron from the adatom escapes directly into an empty bulk state of the underlying material due to the strong surface electric field. The electron effectively tunnels through the energy barrier without the need for additional photon energy. This mechanism is purely field-driven and is more likely when the electric field at the surface is sufficiently high. *(ii)* Photon-

assisted tunneling into the conduction band: here, a photon from a laser pulse excites the adatom electron, increasing its energy. This energy boost allows the electron to tunnel into the conduction band of the material, which may not have been accessible through direct tunneling alone. The photon provides the necessary energy to overcome part of the potential barrier, while the electric field completes the tunneling process. (iii) Photon-induced creation of valence-band holes: in this mechanism, a photon excites an electron from the valence band of the substrate into a higher energy state, leaving behind a hole in the valence band. This hole provides an available state for the adatom electron to tunnel into, facilitating its emission. Unlike mechanism (ii), the photon does not directly excite the adatom electron but rather modifies the electronic states of the substrate to create a pathway for tunneling.

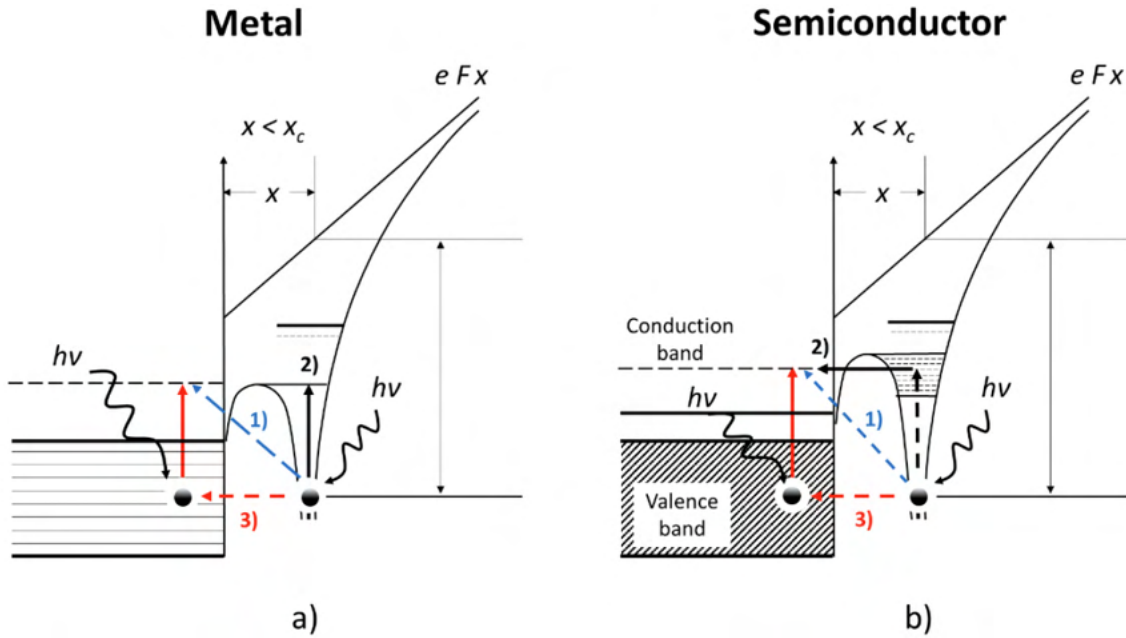


Figure 1.6: The photo-ionization mechanism in a metal (a) or a semi-conductor (b) under an electric field. Arrows and numbers indicate electron excitation (red), relaxation (blue), and tunneling (black) pathways. Adapted from [67].

In metals (Fig. 1.6(a)), the high density of electronic states at the Fermi level and the large electron mobility enable extremely fast charge transport. As a consequence, any positive charge created at the surface by photo-ionization is rapidly screened and neutralized by electrons from the bulk, on timescales much shorter than those required for ion desorption. This efficient re-neutralization suppresses the stabilization of surface ions, rendering photo-ionization experimentally unobservable under typical conditions [68].

In contrast, in semiconductors or other band-gap materials (Fig. 1.6(b)), the presence of an

energy band gap fundamentally alters the charge dynamics at the surface. Under the strong electric fields during field evaporation, significant band bending occurs, leading to the formation of depletion or accumulation regions near the surface. These space-charge regions limit the supply of free carriers and suppress rapid electron–hole recombination. As a result, ions created by photo-ionization can remain stabilized at the surface for a sufficiently long time to undergo field evaporation [67, 69]. For completeness, band bending is the change in the energy of the electronic bands near the surface or an interface of a material, caused by electric fields or charge redistribution.

Overall, photo-ionization can occur either directly through field-induced tunneling or indirectly with the assistance of photon absorption, which either energizes the adatom electron or modifies the electronic structure of the substrate. Metals generally suppress photo-ionization due to rapid electron recombination, whereas semiconductors and insulators, with their finite band gaps and slower recombination rates, allow the ionized atom to persist long enough for field evaporation [69].

From a kinetic perspective, photo-ionization is nearly instantaneous, while field evaporation typically occurs on a sub-picosecond timescale (0.5 ps) [70, 71]. Within this framework, atoms are first ionized to the $1+$ state and can then undergo post-ionization near the tip, consistent with the Kingham theory [72]. Experimentally, photon-assisted evaporation of silicon (semiconductor) has been observed at moderate fields using nanosecond pulses [66] and at low laser energies using sub-picosecond pulses [73].

1.4.2 Thermal Field Evaporation

A short laser pulse can trigger field evaporation by depositing optical energy at the tip apex. This energy is first absorbed by the electronic system and then rapidly transferred to the lattice through electron–phonon coupling within a few picoseconds, producing a transient temperature rise known as the thermal pulse. The evaporation sequence and associated time–temperature window are schematically illustrated in Fig. 1.7. Early studies demonstrated that the apex temperature can rise by several hundred kelvin after a pulse, sufficient to thermally activate field evaporation [74]. As a result, because ions can be emitted throughout the entire cooling period, the spread in emission times broadens the time-of-flight distribution, which can degrade mass resolution. In this regime, the instantaneous evaporation rate follows an Arrhenius-type law is:

$$\phi(t) = \nu \exp \left[- \frac{Q_n}{k_B T(t)} \right], \quad (1.15)$$

where ν is the attempt frequency, Q_n is the activation energy for evaporation, and $T(t)$ is the time-dependent apex temperature. For a wire-like tip geometry (1D heat diffusion), the apex

temperature relaxes according to:

$$T(t, z = 0) = T_0 + \frac{T_{\text{rise}}}{1 + 2t/\tau_{\text{cooling}}}, \quad \tau_{\text{cooling}} = \frac{w^2}{\alpha}, \quad (1.16)$$

where T_0 is the base temperature, α is the thermal diffusivity, w is the effective (Gaussian) half-width of the heated region, and T_{rise} is the initial apex temperature increase. Immediately after electron–phonon thermalization, at $t = 0$ [75], the axial temperature profile is often approximated as a Gaussian:

$$T(t = 0, z) = T_0 + T_{\text{rise}} \exp\left(-\frac{z^2}{2w^2}\right). \quad (1.17)$$

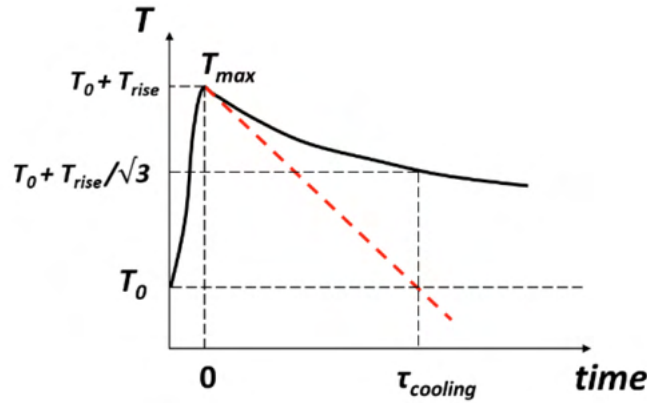


Figure 1.7: The thermal evaporation mechanism: a laser pulse produces a transient temperature rise at the apex followed by diffusive cooling that defines the evaporation time window (black line). The red dashed line represents the normal decrement of the local temperature. Adapted from [23].

For metallic specimens, minimizing the magnitude and duration of the thermal pulse improves mass resolution by narrowing the evaporation time window. This can be achieved by optimizing illumination parameters (shorter wavelength, tighter focus, and shorter pulse duration), adjusting tip geometry for better heat dissipation (e.g., larger cone angles), and lowering the base temperature. In particular, shorter excitation wavelengths reduce the optical absorption depth and effective heated width w , resulting in sharper peaks in the APT spectra [76]. Additional discussions of tip radius and cone-angle effects are given in [23, 70, 76, 77]. In semiconductors and dielectrics, both wavelength and laser fluence strongly influence the peak width and the long thermal tail, representing slow evaporation during cooling.

A typical effect of thermal evaporation that is observed in APT spectra is the presence of

a local minimum following the main peak, succeeded by a delayed secondary maximum. This is typically consistent with a two-step thermal response: an initial rapid temperature rise at the apex due to direct laser absorption, followed by delayed heat flux from the shank toward the tip. This secondary thermal contribution reactivates evaporation after a transient cooling period, producing the observed secondary maximum in the time-of-flight distribution [78, 79]. Finally, defects introduced during FIB preparation, combined with the applied DC field, can further modify absorption and transient heating, as observed in laser-assisted responses [80, 81].

1.5 Atom Probe Tomography applied on biological materials

This section introduces a novel application of Atom Probe Tomography to the determination of protein structures, focusing on the advantages and caveats that such procedure displays in biological contexts.

1.5.1 The Relevance of the Sample Preparation Protocol in Atom Probe Tomography applied on Biology

Initially, APT was used almost exclusively for analysis of metals and alloys but its application range widened in the last decade. To open its usage to low-conductive materials and biomaterials, nowadays a pulsed UV laser illumination is used to control the field evaporation process. APT still presents limitations, as the strong electric field around the needle often induces large mechanical stresses in the sample, called Maxwell stresses. As a result, an APT experiment typically ends with mechanical fracture or field-induced deformation of the specimen, inducing a loss of spatial and chemical resolution.

Up to now APT was discussed as a technique able to evaporate ions from a given (relatively) homogeneous material, a metal or semiconductor (like Silicon). However, its application on macromolecules presents several challenges, due to the intrinsic complex nature of biological objects. In fact, the construction of a needle-shaped specimen made of only proteins is not feasible. Thus, this specific APT workflow must somehow include a preliminary step-0 protocol able to generate a thin needle containing the biological sample, protecting it and avoiding damage or fracture during the subsequent APT analysis. In fact, such needle must withstand cryogenic temperatures, ultra-high vacuum, and intense electric fields, experimental conditions largely incompatible with the physiological environment of biomolecules in solution. Therefore, the design of such protocol represents the biggest obstacle to surmount in this context, as well as its reliability and reproducibility.

1.5.2 The First Application of Atom Probe Tomography on Proteins

Recently, a novel experimental workflow for investigating the 3D structure and chemical composition of individual proteins using APT was introduced by a research group at Chalmers University of Technology in Gothenburg, Sweden [82]. This approach extends the applicability of the method beyond traditional materials science and opens new possibilities for structural biology, particularly for proteins. In fact, Sundell *et al.* [82] demonstrated that APT might have the potential to reconstruct single protein structures in 3D without relying on ensemble averaging, marking a fundamental paradigm shift in structural biology. Rather than deriving a consensus structure from millions of particles, APT enables true single-molecule 3D reconstruction, preserving structural individuality and molecular heterogeneity. Crucially, APT couples to spatial coordinates (X, Y, Z) also the chemical sensitivity through mass-to-charge identification, as aforementioned in Section 1.2. This capability allows direct determination of elemental composition at near-atomic resolution, revealing chemical heterogeneity and local compositional variations within a single biomolecule. The relevance of the APT process therefore lies not only in its spatial resolving power, but in its unique integration of structural and chemical information, enabling a more complete and molecule-specific understanding of protein architecture and function, at the level of individual molecules.

The proposed method builds upon earlier specimen preparation strategies for APT [83], with specific modifications designed to overcome the intrinsic challenges associated with biological materials. This novel sample preparation protocol and its details are reviewed in Section 1.5.3. In particular, the aqueous environment surrounding the protein should be by an amorphous silica matrix formed *via* a tailored sol–gel process. In principle, this gelled silica network should provide mechanical stability and protects the embedded biomolecule from damage during focused ion beam (FIB) milling and the following APT analysis, while preserving the native structure of the protein. Importantly, the proposed encapsulation process can be carried out at room temperature. In addition, silica has previously been shown to partially maintain embedded enzymes in a functional state [84].

Fig. 1.8 (a) illustrates the gel-like silica matrix formed from a silica sol containing the protein of interest. The needle-shaped APT specimen shown in Fig. 1.8 (b) was fabricated using an *in situ* FIB–SEM lift-out procedure [85], followed by a standard tip-sharpening step (commonly employed in APT), also described in Section 1.5.3. The resulting silica matrix should confer sufficient mechanical integrity to the specimen, enabling controlled APT analysis without premature fracture [84].

In this Chalmers study, rabbit immunoglobulin G (IgG) was selected as a model protein [86]. Successful silica encapsulation requires a careful balance between maintaining a chemically benign environment for the protein and forming a sufficiently dense and robust silica matrix.

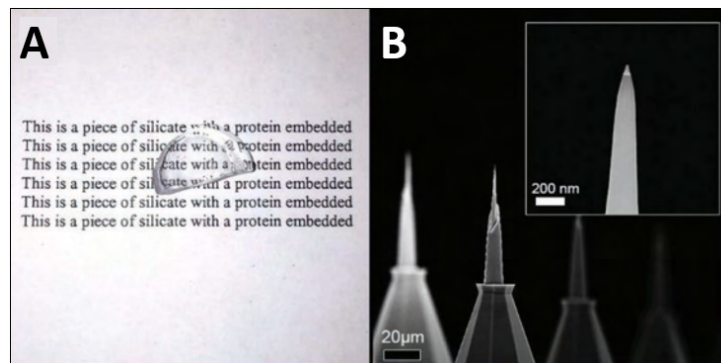


Figure 1.8: Gel-like silica matrix embedding the proteins before APT analysis. (a) Macro photograph of a bulk piece of the synthesized protein-containing silica matrix. (b) SEM micrograph (600X magnification) of an array of protein-containing specimen tips sharpened using FIB. Insert: SEM micrograph of individual specimen apex at 100 000X magnification. Adapted from [82].

Since different proteins respond differently to pH and ionic conditions, multiple synthesis routes were explored. Two distinct silica precursors were investigated: tetraethyl orthosilicate (TEOS) and sodium silicate (commonly referred to as water glass), yielding two amorphous silica matrices with markedly different properties.

Silicon alkoxides such as TEOS undergo spontaneous hydrolysis upon contact with water, followed by condensation reactions that form an amorphous silicon–oxygen network while releasing alcohol as a by-product. These reactions are catalyzed under both acidic and basic conditions and are strongly influenced by parameters such as pH, precursor concentration, temperature, and the nature of the organic substituents [87]. By adjusting these parameters, the gelation time and physical properties of the silica matrix, including porosity and density, can be controlled. Under neutral pH conditions, TEOS-derived silica was found to be excessively porous and mechanically unstable for APT analysis. In contrast, synthesis under acidic conditions produced a dense silica matrix capable of withstanding the mechanical stresses encountered during APT. However, IgG embedded in low-pH TEOS-derived silica exhibited pronounced aggregation, likely due to the sensitivity of the F_c domain to acidic environments, as shown in Fig. 1.9 (a). Similar F_c -mediated aggregation has been reported to promote complement activation [88].

More favorable results were obtained using sodium silicate as the silica precursor. In aqueous solution, the simplest silica species is orthosilicic acid, $\text{Si}(\text{OH})_4$, which readily polymerizes at concentrations exceeding approximately 10^{-3} M, forming nanoscale aggregates that ultimately condense into an amorphous solid as water evaporates [89]. Because sodium silicate solutions are inherently basic, an ion-exchange step was employed to adjust the pH to near-physiological values prior to protein addition. Under these conditions, no IgG aggregation was observed, and

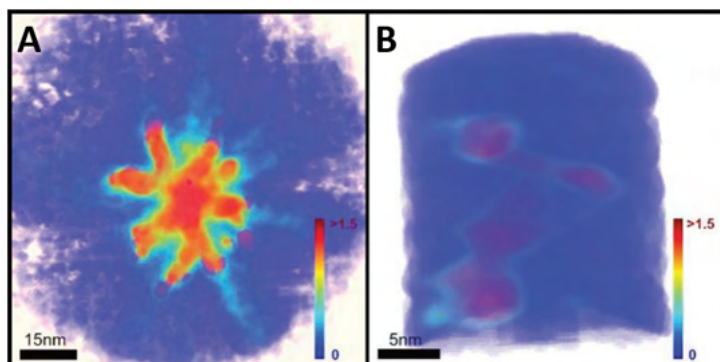


Figure 1.9: (a) Pentameric aggregate of IgG is visible in a 2D heatmap (atomic number density of organic species) from an atom probe analysis with TEOS-derived silica at pH 2 as the embedding matrix. (b) Heatmap (atomic number density of organic species) of a single molecule of IgG embedded in a silica matrix composed of orthosilicic acid ($\text{Si}(\text{OH})_4$), synthesized through auto-polycondensation at pH 7. Adapted from [82].

a two-dimensional APT reconstruction of a single IgG molecule is shown in Fig. 1.9 (b).

Electron-transparent lamellae prepared from the $\text{Si}(\text{OH})_4$ -derived silica matrix using FIB–SEM confirmed, *via* bright-field transmission electron microscopy, the presence of an intact IgG monomer exhibiting the characteristic Y-shaped morphology (Fig. 1.10 (a)). Since the silica matrix is inorganic, reconstruction of the protein relied on carbon-containing ionic species, allowing for an unambiguous spatial mapping of a single IgG antibody. The distribution of such species is presented in Fig. 1.10 (b). Due to the intrinsic flexibility of the antibody hinge regions, complete crystallographic structures of IgG molecules are scarce. Consequently, the reconstructed rabbit IgG was compared with a known human IgG structure obtained by X-ray diffraction at 2.7 Å resolution (PDB entry 1HZH) [90]. This comparison, illustrated in Fig. 1.10 (c), was performed by importing the crystallographic model into atom probe reconstruction software and generating 3D atomic density maps for both structures.

The spatial resolution of APT is anisotropic, typically being superior along the analysis direction compared to lateral directions. As a result, assigning a single global resolution value is challenging. Nevertheless, the clear visibility of F_{ab} loop features suggests a resolution better than 10–15 Å on a single IgG molecule, sufficient to resolve tertiary structural elements, although not fine details such as secondary structure or individual side-chain rearrangements.

The atom probe mass spectra further corroborate the successful encapsulation of IgG within the silica matrix (Fig. 1.11). Due to the fundamentally different ionization mechanisms involved, APT spectra of biomolecules differ significantly from those obtained using conventional mass spectrometry techniques. The dominant organic ions observed were CNH^{2+} , likely originating

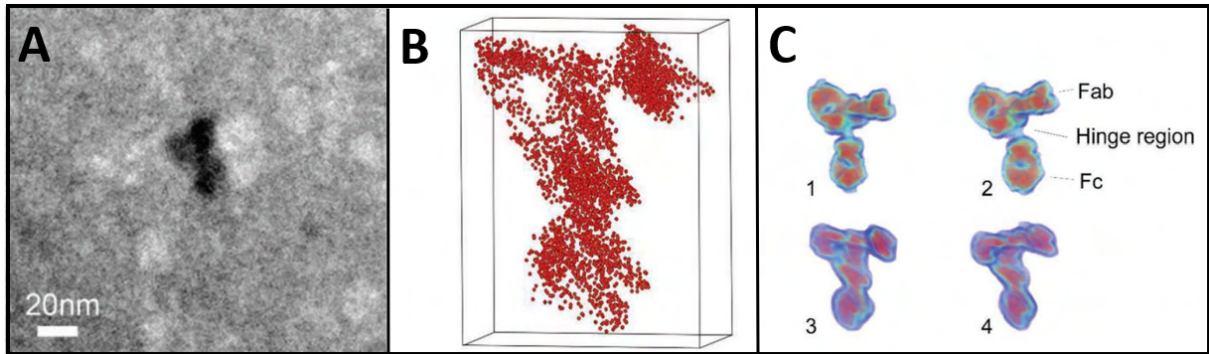


Figure 1.10: TEM micrograph and atom probe 3D reconstructions. (a) Bright field TEM micrograph of a single IgG molecule embedded in Si(OH)_4 -derived silica matrix. (b) Spatial distribution of aminoacidic ionic species CNH_2^+ and CO_2^+ in a single IgG subvolume, retrieved by generating an isodensity region with the more than 6 C atoms nm^{-3} . (c) the images in numbers 1, 2 represent a carbon isodensity heat map (5 C atoms nm^{-3}) of human IgG as reported by Saphire *et al.* (2001) [90] and regenerated in the atom probe 3D visualization software. 3 and 4 display the equivalent isodensity surface from the APT analysis, retrieved from the isodensity regions. To assemble the acquired data into tomograms, a standard voltage evolution-based reconstruction protocol for APT was adopted [26]. Adapted from [82].

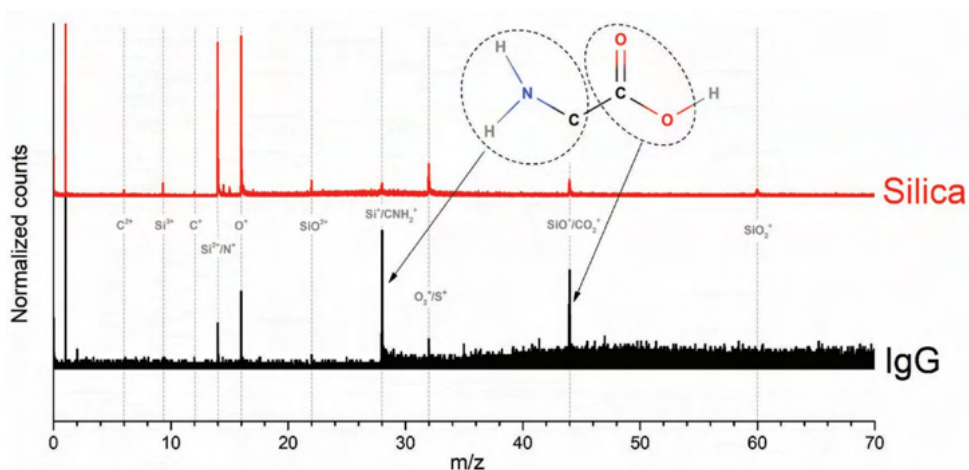


Figure 1.11: Atom probe mass spectra. Annotated mass spectra in linear scale from the Si(OH)_4 -derived silica matrix (red) and the IgG subvolume (black) of an atom probe analysis. The peaks from amino acid trace ions are indicated with arrows.. Adapted from [82].

from combined amine and α -carbon fragments, and CO^{2+} , associated with carboxyl groups. Overlapping peaks between the protein and silica matrix complicate the identification of certain elements, such as nitrogen and sulfur, necessitating isotopic labeling for unambiguous assignment [91]. Quantitative analysis was therefore limited primarily to carbon. After correcting for detection efficiency and isotopic abundance, the measured carbon content was found to be in good agreement with reported values for rabbit IgG [86].

Notably, characteristic water-related peaks at 17–19 Da were absent from the spectra, suggesting that the hydration shell surrounding the protein was largely replaced by silica during matrix formation. Nevertheless, the presence of oxygen and hydrogen ions throughout the analysis indicates that residual water cannot be entirely excluded, as also stated in other experimental works [91].

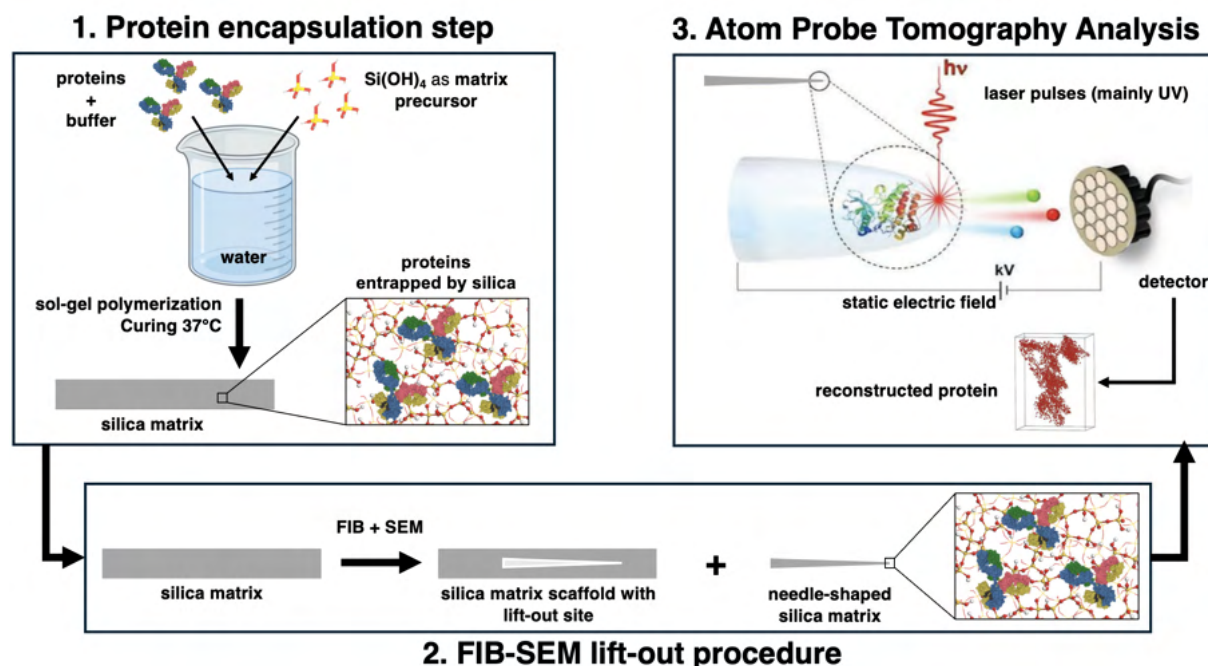


Figure 1.12: Workflow of the First Atom Probe Tomography Analysis on Proteins, consisting of three key steps: protein encapsulation (1), FIB-SEM lift-out procedure (2) and APT analysis itself (3). Adapted from [82].

A complete workflow of this application of APT on the determination of protein structures is reported in Fig. 1.12. In summary, the work by Sundell *et al.* [82] demonstrates a fundamentally new approach for protein analysis using APT. With continued methodological development, this technique holds promise for simultaneously providing 3D structural information of single protein molecules flanked with elemental composition, achieving what stated in Section 1.2. Tor

recall, this novel protocol could have the potential to yield an analysis of individual proteins, so in principle guaranteeing access to the structural conformation of each protein unit encapsulated in the silica matrix. While atomic-scale resolution comparable to crystalline materials is unlikely for biomolecules, the method already enables access to protein tertiary structure and, potentially, conformational variability. The mechanical robustness of the $\text{Si}(\text{OH})_4$ -derived silica matrix further supports its suitability as a matrix precursor for biological APT studies.

1.5.3 Details of the Sample Preparation Protocol and the FIB-SEM Procedure

As reported in Fig. 1.12, the overall procedure can be divided into a sequence of well-defined steps: (i) sample preparation protocol (protein encapsulation), (ii) FIB-SEM lift-out procedure, (iii) APT analysis itself. The current section focuses on the first two steps of this pipeline proposed by Sundell *et al.* [82], and the last will be discussed in Section 1.5.4.

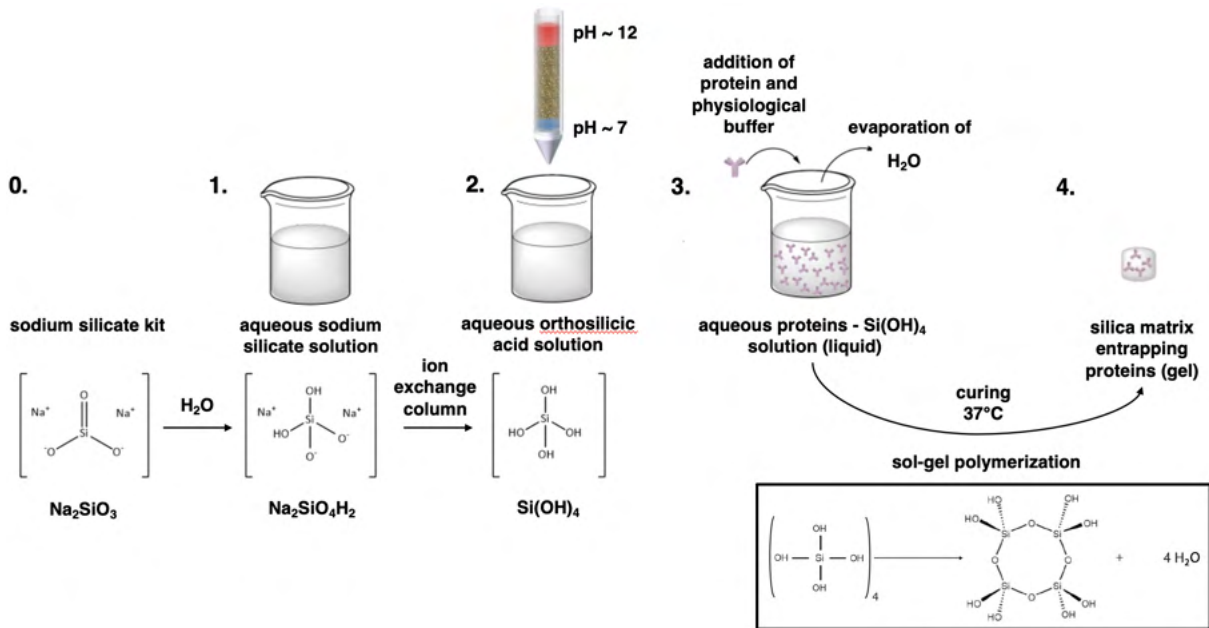


Figure 1.13: Scheme of the Sample Preparation Protocol. Sodium silicate (step 0.) is inserted into an aqueous solution (step 1.), and then passed into an ion exchange column, obtaining an aqueous $\text{Si}(\text{OH})_4$ solution (step 2.). Once reached pH values close to 7, proteins are added (step 3.), and the whole mixture undergoes a sol-gel polymerization step, passing from a liquid to a gel-like state (step 4.).

The preparation of a liquid solution containing both proteins and the matrix precursor molecule, here $\text{Si}(\text{OH})_4$, starts generally from commercially available kits, like sodium silicate kits (step

0. in Figure 1.13). The matrix primer kit might also be more complex in terms of composition, comprising a monomer, a crosslinker, and an initiator. The initiator provides the energy required to trigger polymerization, leading to the formation of a matrix through bonding between monomer and crosslinker. Consequently, the transition from the liquid state to the gel-like state depends strongly on the chosen matrix precursor. For polymer-based matrices, a curing phase might be needed at higher temperatures (like 37 °C), to supply sufficient energy to accelerate the matrix formation.

The highly alkaline sodium silicate Na_2SiO_3 (pH 11-12) is dissolved in water (step 1. in Figure 1.13), meaning that silica exists mainly as silicate ions SiO_3^{2-} and not in the form of $\text{Si}(\text{OH})_4$. Then, the whole mixture is passed through an ion exchange column (step 2.), activated with hydrochloric acid (HCl) to reduce the pH of the solution up to 7.5 circa. The chemical reaction characterizing this process is the following: $\text{Na}_2\text{SiO}_3 (\text{aq}) + 2 \text{HCl} (\text{HPLC/resin}) \longrightarrow \text{Si}(\text{OH})_4 (\text{aq}) + 2 \text{NaCl} (\text{aq})$. Such reaction reduces the pH from highly alkaline to mildly basic/neutral (7-7.5), as silica begins to be mostly present as an acid, $\text{Si}(\text{OH})_4$, reaching metastability. Intuitively, the role of the pH in this whole procedure is extremely relevant. Briefly, the kinetics of the $\text{Si}(\text{OH})_4$ polymerization is highly dependent on pH of the solution. $\text{Si}(\text{OH})_4$ is a weak acid, and its $\text{pK}_{\text{a}1}$ is close to 9.5. This means that at pH close to 9.5 the concentration of $\text{Si}(\text{OH})_4$ and the conjugated base (or anionic form, $\text{Si}(\text{OH})_3\text{O}^-$) is equal, i.e. half of the monomers are with/without a proton. For pH values lower than $\text{pK}_{\text{a}1}$, $\text{Si}(\text{OH})_4$ is the most abundant species, whereas for higher pH values the deprotonated form $\text{Si}(\text{OH})_3\text{O}^-$ becomes dominant. Intuitively, the polymerization rate finds its sweet spot at low values of pH, where there is lots of substrate (i. e. unprotonated $\text{Si}(\text{OH})_4$) and few $\text{Si}(\text{OH})_3\text{O}^-$ anions; in this condition polymerization occurs effectively. However, too low pH values lead to insignificant concentrations of the anions (all monomers are protonated), and no polymerization occurs. Contrary, for high pH values, the high negative charge accumulated by the $\text{Si}(\text{OH})_3\text{O}^-$ molecules generates repulsion among monomers, dampening the polymerization rate.

Once reached pH values close to 7/7.5, the proteins and the relative buffer are added (Figure 1.13, step 3.). At such pH value, the polymerization rate is not at its maximum, but is still sufficient to guarantee a good formation of the silica matrix. In the specific case of orthosilicic acid, auto-condensation can also occur even at room temperature at sufficiently high concentrations (greater than 10^{-3} M), albeit with relatively slow kinetics. For this reason, the initial employed $\text{Si}(\text{OH})_4$ concentration ranges between 0.3 and 0.6 M. Lastly, the whole mixture is cured at 37°C, waiting for silica to polymerize, forming a connected amorphous gel-like matrix (step 4.). Overall, external control parameters are therefore often required to promote efficient condensation. Depending on variables such as concentration, pH, temperature, and reaction time, the silica network may adopt different configurations, including chain-like, ring-like, or

branched architectures, which eventually reorganize into a continuous, amorphous gel structure.

This whole procedure is characterized by one key and essential phase: the silica should progressively replace the native aqueous environment around the protein structure during the early stages of the $\text{Si}(\text{OH})_4$ polymerization, as the presence of water between the biomolecules and the matrix during the subsequent APT analysis would definitely lead to failure. Moreover, at a certain stage (see Figure 1.13, step 3), a significant amount of water is removed to accelerate the polymerization chain reaction. If the proteins have not interacted with the silica beforehand, they will be removed together with H_2O .

Previous studies have demonstrated that enzymes encapsulated in silica matrices can partially retain their catalytic activity and, in few cases, remain almost unperturbed, as evidenced by solid-state NMR investigations [92]. However, it is reasonable to expect at least minor conformational changes of the protein due to the interaction with a different chemical environment, at least at the level of amino acid side chains.

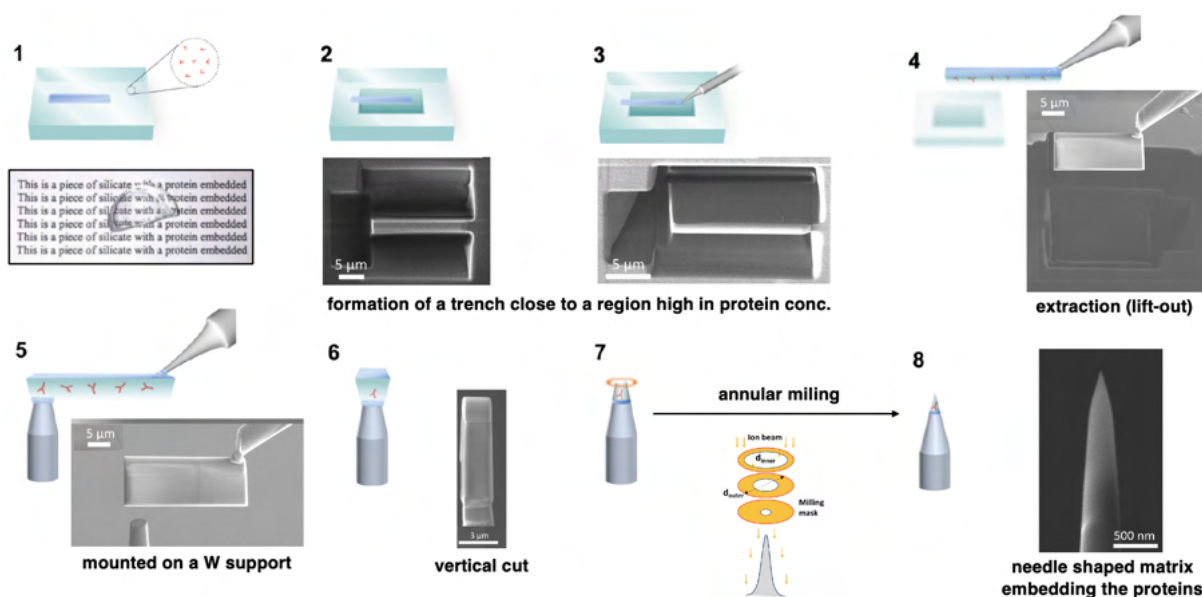


Figure 1.14: Scheme of the FIB-SEM Procedure. Schematic representation of the APT lift-out process used to obtain specimens from a silica sample containing the embedded proteins. A region high in protein concentration is identified and the FIB is used to lift out (1-4) such volume, placing on flat top posts (5). The lift-out segments were then milled using the ion beam into APT ready specimens with diameters below 100 nm (6-8).

To conclude, the success of the atom probe tomography (APT) pipeline for protein analysis depends critically on the sample preparation step, and specifically from its early stages. Effective encapsulation of proteins within an amorphous silica matrix is essential to preserve their spatial

integrity throughout dehydration, focused ion beam (FIB) processing, and field evaporation. The sol–gel-derived silica matrix provide sufficient mechanical stability for specimen fabrication and analysis. At the same time, it must remain chemically inert and structurally homogeneous to avoid obscuring the protein signal during laser-assisted evaporation. Inadequate control over this preparation step can lead to protein displacement, damage, or loss of contrast relative to the surrounding matrix, thereby limiting the achievable spatial and chemical resolution in APT reconstructions.

Figure 1.14 reports the FIB-SEM procedure. Once gelation is achieved, the bulk material is mechanically cleaved (steps 1–3) with the FIB-SEM *via* a methodology called the wall method, an operation exposes the region of the matrix containing the highest concentration of proteins, which typically settles toward the bottom of the container during preparation. Step 4 depicts the extraction of the protein-rich region from the silica matrix. All these steps have a peculiar, essential purpose: to reduce the macroscopic dimensions of the sample, which are incompatible with APT analysis. This silica region high in protein concentration is then mounted on a top post, glued and reduced (again) in dimension (step 5-6).

The extracted volume is subsequently shaped into a needle-like geometry through an annular milling procedure of diameters below 100 nm, as illustrated in steps 7-8. This particular geometry is a fundamental requirement for APT, as described in detail in standard APT literature [21] and also in Chapter 4, Section 4.1.3. The final needle-shaped specimen is then suitable for APT analysis.

1.5.4 Details of the Laser-Assisted APT Analysis

This Section has the objective to elucidate some details about the third main step of the protocol, the la-APT analysis, focusing mainly on its crucial and critical features.

As stated in Section 1.2, APT instruments can resolve isotopes: such aspect is extremely relevant for this pipeline characterized by hybrid organic–inorganic systems such as IgGs embedded in silica matrices. In fact, a central limitation of APT is the occurrence of ambiguities in the mass-to-charge (m/z) spectrum, arising from carbon-, nitrogen-, and oxygen-containing ionic species deriving from the IgGs, that produce indistinguishable peaks (i.e. overlaps). This issue becomes more severe in silica–protein systems due to the coexistence of such light elements (H, C, N, O) and silicon, combined with the extensive fragmentation that occurs during field evaporation. As a consequence, distinguishing chemical species based solely on m/z becomes highly uncertain. For example, carbon ions (C^+ , 12 Da) and nitrogen-related fragments such as N^+ and CH_2^+ (both 14 Da) could coincide. Similarly, IgG generates many molecular ions such as CO^+ and N_2^+ , both appearing at 28 Da, which coincides with the peak of the most abundant silicon isotope, ^{28}Si . Such overlap among the three species introduces significant noise, making it

difficult to unambiguously assign the detected signal to either the silica matrix or the embedded biomolecule. Additional complexity arises from oxygen- and hydrogen-containing species. Ions such as O^+ (16 Da), OH^+ (17 Da), and H_2O^+ (18 Da) can originate from both the sample and residual water, and often overlap with hydrocarbon fragments. Hydrogen ions (H^+ , H_2^+) are also ubiquitous and provide limited chemical specificity due to their multiple possible sources.

For these reasons, a powerful strategy to mitigate this issue is isotopic labeling, which shifts characteristic peaks by replacing naturally abundant isotopes with less common ones. For instance, ^{13}C and ^{15}N labeling shifts protein-related peaks to higher m/z values. This reduces peak overlap and enables a clearer distinction between matrix- and protein-derived ions. Sundell *et al.* [82] utilized engineered ^{13}C -labeled IgG molecules.

Another relevant aspect lies in the possible critical aspects that might arise in reconstruction phases of specimens containing silica and biomolecules. As stated in Section 1.3.1, equation 1.18 constitutes a known relationship between voltage, evaporation field and tip radius that can be utilized to get spacial information of homogeneous, conductive materials, as the evaporation field (and thus the applied voltage) can be assumed to be relatively constant across the specimen. Briefly, the voltage can be utilized to track the evolution of the tip radius and refine the signal of the depth (z-coordinate) of evaporated atoms. This means that, for every atomic species, equation 1.18 can become:

$$F_{evap} = \frac{V}{k_F R}, \quad (1.18)$$

where F_{evap} identifies an evaporation threshold typical of each element evaporated. If the specimen is homogeneous, every atom will share the same F_{evap} . However, using such relation $F = V/k_F R$ in heterogeneous systems like silica-protein systems becomes critical for several reasons. To get this concept, it is first necessary to explain a relevant concept, namely the laser pulse energy (LPE).

Under laser-assisted APT conditions, evaporation is not purely field-driven but also strongly influenced also by thermal effects (see Section 1.4.2). In this regime, LPE plays a central role in determining the evaporation behavior, as the effective evaporation field threshold (or barrier) is reduced by a local temperature increase induced by the LPE, as also shown in equations 1.11 and 1.14 of Section 1.3.3. As consequence, the final condition for evaporation becomes a coupled function –a summation– of both electric field (voltage) and temperature (or LPE) $F_{evap}(V, T)$, rather than a purely electrostatic threshold (F_{evap}). This means that, the interplay between these two parameters, voltage and LPE, is thus extremely crucial. In other words, a fixed target final evaporation rate, can be described by a high field contribution and a low LPE contribution, or vice versa. This two combined contributions the actual local evaporation behavior: as far as the material is homogeneous, this is not a problem, but if we consider a specimen of silica and embedded IgGs, things change. In fact, the two entities exhibit markedly different optical

absorption, thermal conductivity, and heat dissipation properties. As a result, a given final evaporation rate $F_{evap}(V, T)$ does not produce a uniform evaporation condition, but a spatially heterogeneous one. Typically, IgG regions characterized by light atoms like ^{13}C , N, O display lower electric field evaporation F_{evap} thresholds with respect to silica (inorganic, heavier).

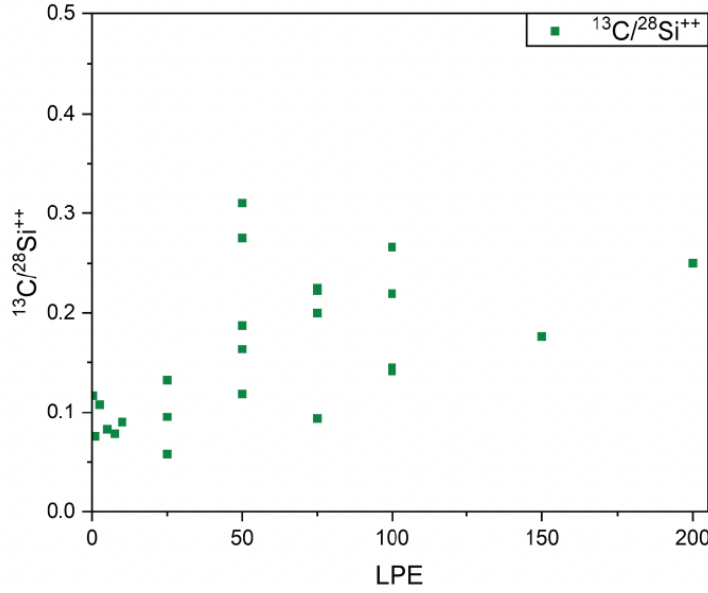


Figure 1.15: The measured $^{13}\text{C}/^{28}\text{Si}^{2+}$ ratios at various LPEs [pJ] from different analyses where the $^{13}\text{C}/^{12}\text{C}$ ratio was above 5. Given that the evaporation rate in la-APT is always constant, for low LPE values (i.e. high voltage values) both the protein and the matrix evaporate, but $^{13}\text{C}/^{28}\text{Si}^{2+}$ ratio is low because ^{13}C atoms cannot be efficiently detected. Instead, for high LPE values (i.e. low voltage values), the ratio is higher, but the total reconstruction efficiency was low (not shown). Close to middle-range values of LPE, the detection of ^{13}C is optimal (the ratio is high), as both the matrix and the IgG exhibit a similar evaporation trend, leading to good final reconstruction.

At low voltage/electric fields (meaning high LPE, if the final evaporation rate is kept constant), protein atoms evaporate, but silica has not reached its evaporation threshold yet. This nonuniform evaporation phenomenon has bad consequences on the overall final result: since IgG atoms are trapped within a slowly-evaporating silica, they cannot immediately leave the tip, displaying surface diffusion and subsequently directional migration toward the outer regions of the needle-shaped specimen, ultimately leaving it. This causes negative effects during reconstruction, as APT assumes that the evaporation occurs uniformly layer-by-layer: if ^{13}C exhibits migration toward outer layers and subsequently evaporates, it gets assigned to the same depth and radius of curvature of silica atoms in those outer layers, leading to artifacts in the final structural profile of the tip/biomolecule. Instead, at high voltage/electric fields (meaning low LPE, if the

final evaporation rate is kept constant), silica efficiently evaporates, but given the net difference between the number of atoms coming from the matrix and the IgGs, protein information is lost across the analysis (its evaporated ions are not captured as detectable by the instrument). In conclusion, performing APT on such a heterogeneous specimen requires identifying an optimal balance between voltage and LPE that yields a sufficiently uniform final evaporation rate, as shown in Fig. 1.15. Achieving this compromise is essential to minimize differential evaporation between components and to ensure a reliable reconstruction of all regions of the tip, including both the silica matrix and the embedded protein.

To recall, the applied static voltage is no longer uniquely related to the final evaporation field (laser + voltage). This breaks the fundamental assumption that voltage alone can be used to track the spatial evolution of the tip radius corresponding to a unique geometrical state of the tip. Therefore, voltage cannot be reliably used as a reconstruction parameter in laser-assisted APT of heterogeneous systems. Instead, reconstruction strategies must rely on geometrical models, ion detection sequences, and assumptions about volume conservation.

1.6 The MIMOSA Project

Despite the promising results reported by Sundell *et al.* [82], several fundamental questions regarding the proposed protocol remain unresolved. More generally, while the work of Sundell *et al.* represents a major step forward, it also highlights several limitations that currently hinder atomic-scale analysis of biological materials using APT. These include insufficient chemical resolution to unambiguously resolve the molecular structure of proteins, a low signal-to-noise ratio that complicates reliable 3D reconstruction, and uncertainties in spatial resolution arising from specimen heating induced by laser-assisted APT operation.

In an effort to address these challenges, a European research project entitled 4D Microscopy of biological materials by short pulse terahertz sources (**MIMOSA**) was launched on 1 September 2022. The primary objective of MIMOSA is to bridge the gap between current industrial and academic developments in APT imaging. MIMOSA therefore aims to develop a prototype microscope capable of providing simultaneous spatial and chemical characterization of biological materials at the atomic scale, invoking the concepts already mentioned in Section 1.2. Thus, the final objective of this project is to complement results from cryo-EM/X-ray crystallography, that otherwise suffer from intrinsic limitations when used independently.

This new instrument concept seeks to utilize new laser types during APT analysis, that might display a better fit with biological materials. In recent years, the high-resolution microscopy community has actively explored alternative excitation schemes to overcome the limitations of conventional laser-assisted APT. A particularly innovative approach was introduced by Vella *et*

al. [93], who demonstrated the feasibility of controlling ion emission from a metallic nanotip using intense THz pulses, instead of standard ultraviolet (UV) pulses. This work established new degrees of freedom for manipulating the atomic evaporation process and represents a significant advance in atom probe instrumentation.

The interaction between intense THz pulses and biocompatible materials, as biomolecules embedded in silica matrices, remains largely unexplored. While the mechanisms of UV, visible, and infrared laser-assisted field evaporation have been extensively studied, the response of materials to THz excitation is still poorly understood. A key advantage of THz-driven evaporation lies in its potential to enable non-thermal field evaporation (see Chapter 4, Section 4.1.2 for more details). This is particularly relevant for soft organic materials such as proteins, which are highly sensitive to temperature-induced damage. By minimizing thermal effects, THz-assisted APT could help ensure that biomolecules retain their native structure throughout the analysis.

In parallel, advances in detector technology further enhance the prospects of atom probe microscopy. State-of-the-art APT detectors already achieve detection efficiencies exceeding 80%, and realistic pathways exist for the development of new ion detection concepts incorporating kinetic energy discrimination [94, 95]. If successfully implemented, such advances could enable atom probe analyses approaching high detection efficiency, allowing the precise localization of every atom within a single protein without the chemical ambiguities caused by spectral overlaps.

The sample preparation protocol is another procedure that MIMOSA aims to largely improve. Silica matrices could be derived from a broader range of molecular primers, each imparting distinct chemical and structural features to the resulting network. Beyond tetraethyl orthosilicate (TEOS), alternative primers derived from $\text{Si}(\text{OH})_4$ might offer the potential to tune matrix polarity, porosity, and interfacial interactions with embedded biomolecules. Such primers may enable the incorporation and stabilization of more complex targets, including membrane proteins, by promoting favorable interactions between hydrophobic domains and the evolving silica framework. By tailoring the primer chemistry, it may be possible to generate hybrid silica matrices that better accommodate amphiphilic proteins, preserve their native conformations, and limit aggregation or denaturation during dehydration and field evaporation. Expanding the range of silica precursors therefore represents a promising strategy to extend APT analyses beyond soluble proteins toward membrane-associated systems that remain challenging under conventional preparation protocols.

In addition, further optimization of the sol-gel synthesis protocols will also be required to routinely produce suitable $\text{Si}(\text{OH})_4$ -derived silica matrices. Standard FIB-SEM-based APT specimen preparation typically requires volumes on the order of 50 fL to fabricate several needle-shaped specimens. If biomolecules can be homogeneously dispersed within the silica matrix, low protein concentrations could, in principle, be sufficient for APT-based structural analysis. This

capability would enable seamless integration with lab-on-a-chip platforms, where only minute liquid volumes are available for investigation.

The potential impact of combining THz excitation with matrix functionalization is therefore considerable. Achieving high levels of performance would profoundly enhance our understanding of complex biomolecular systems, including ligand–receptor interactions and metal ion coordination in heme-containing proteins. Similar to the transformative impact of direct electron detectors in cryo-EM, the development of a new THz-assisted atom probe platform could trigger a paradigm shift in nanoscale biological analysis. Beyond fundamental research, such advances may also have far-reaching implications in medicine and pharmaceutical science. More than half of modern therapeutic drugs target proteins, particularly membrane proteins, for which detailed 3D structural information is often lacking. By providing new insights into these systems, MIMOSA has the potential to significantly accelerate drug discovery and development, ultimately yielding substantial societal benefits.

1.7 The Role of *in Silico* Modeling in the MIMOSA Project

As stated in the previous section, acquiring a deeper understanding of the matrix effect on several kinds of proteins and defining also a reliable sample preparation protocol that can be applied to a wide range of biomolecules represent two major objectives of the MIMOSA project. Also the design of new laser types (like THz) able to induce less thermal effects and stresses during the APT analysis are a huge milestone that must be pursued to unlock the full potential of this novel protocol.

Notably, the success of this methodology critically depends on the compatibility between the biomolecule and the embedding silica matrix. The encapsulation must preserve the native structure of the protein, minimize mechanical stress, and ensure a homogeneous and stable coating. These stringent requirements introduce significant uncertainties that cannot be fully addressed by experimental techniques alone. A valuable contribution could come from *in silico* modeling, which represent a complementary approach capable of formulating testable hypotheses and identifying physically meaningful pathways, thereby guiding experimental design and facilitating the interpretation of complex experimental observations. Therefore, the current section is organized into two separate subsections. The former describes the computational strategy adopted to simulate the key stages of the protein sample preparation protocol, while the latter focuses on the modeling approach used to reproduce and interpret the subsequent APT analysis. A much more exhaustive explanations of such computational procedures will be given in Chapters 3, 4, and 5, respectively.

1.7.1 Molecular Mechanics Methods Might Unravel Key Mechanisms of the Sample Preparation Protocol

Immunoglobulin G (IgG) is a relatively large and structurally rigid protein, composed of a compact F_c core and two shorter, more flexible F_{ab} domains. Although according to Sundell *et al.* [82] silica should preserve the IgG conformation, it remains unclear whether the observed compatibility between the protein and the silica matrix is influenced by the size and rigidity of IgG. Nevertheless, these previous investigations have mainly focused on relatively large and robust proteins (like IgG antibodies), able to sustain lots of mechanical stresses. To address this issue, future investigations should focus on smaller and more flexible proteins, whose structural dynamics may be more sensitive to interactions with the silica network already during the early stages of the $\text{Si}(\text{OH})_4$ polymerization. In this context, developing numerical simulations to model protein incorporation into a silica matrix would be highly beneficial. Classical Molecular Dynamics (MD) simulations [96] can provide valuable guidance for experimental procedures, functioning as an *in silico* microscope able to visualize and characterize the interaction protein–silica, in terms of electrostatics, van der Waals forces, hydrogen bonding and hydrophobic effects. MD gives a time-resolved, atomistic description of the system, enabling the explicit modeling of protein–silica interactions in aqueous environments. By integrating accurate force fields for both biomolecules and silica precursors, MD could unravel the dynamical pathways leading to encapsulation. Moreover, the embedding process is inherently dynamic and occurs at nanoscopic scales, involving transient and collective effects that are not directly observable experimentally. This lack of direct experimental access motivates the use of atomistic simulations. More details about the Molecular Dynamics algorithm can be found in Chapter 2, whereas details about how Molecular Mechanics methods can be used to unravel key mechanisms of the sample preparation protocol can be found in Chapter 3.

1.7.2 First Principles Calculations Might Clarify Crucial Aspects of the Ion-Field Evaporation in APT analyses

From a theoretical perspective, the investigation of THz-driven ion field emission requires an explicit treatment of the electronic degrees of freedom [97]. In particular, a time-dependent framework is essential to capture the inherently nonlinear response arising from the simultaneous interaction of the sample with an intense THz excitation and a static field. Such a scenario induces dynamical charge redistribution and time-dependent modifications of interatomic forces, which cannot be described within static or adiabatic approximations.

Time-dependent density functional theory (TDDFT) [98] provides a rigorous quantum-mechanical description of electronic excitations in external fields and can be naturally coupled

to molecular dynamics within the Ehrenfest formalism [99]. This *ab initio* Molecular Dynamics approach enables a self-consistent propagation of electronic and ionic degrees of freedom, making it, in principle, suitable for addressing field-driven evaporation phenomena. Nevertheless, computational investigations of laser- or THz-assisted field evaporation based on TDDFT remain scarce. Existing studies have primarily focused on femtosecond UV laser pulses [100] or on static and quasi-static first-principles approaches to field evaporation [101, 102, 103]. Despite these advances, the evaporation of cations from a solid surface constitutes an exceptionally complex, multiscale process. Although *ab initio* simulations offer a formally rigorous route to its investigation, a fully realistic description is currently beyond reach. This limitation stems from the coexistence of multiple, strongly coupled mechanisms that contribute to bond breaking at the specimen surface during an APT experiment. In particular, electronic ionization and thermal effects act concurrently in promoting atomic evaporation. While ionization processes can be described within first-principles electronic-structure frameworks, thermal contributions involve electron–phonon coupling, lattice heating, and phonon-mediated energy redistribution occurring over longer and frequently stochastic timescales. These dissipative processes are challenging to treat accurately within first principles MD and precisely TDDFT, which remains a mean-field approach and does not provide a fully realistic account of decoherence or temperature-dependent energy dissipation.

For these reasons, a comprehensive and quantitatively predictive simulation of the complete evaporation event lies beyond the present methodological capabilities. Accordingly, the present work concentrates primarily on the role of electronic ionization in the context of APT field evaporation, as this mechanism is both conceptually well defined and directly accessible within an *ab initio* time-dependent framework. A detailed discussion about the *ab initio* Molecular Dynamics procedure employed in this thesis (TDDFT and Ehrenfest integrations schemes) is provided in Chapter 2, Sections 2.4 and 2.5). Lastly, more details about the results that this approach led in the context of APT can be found in Chapter 4.

Foundations of Molecular Mechanics and First-Principles Simulations

This Chapter introduces the basis of two computational methods utilized in this thesis: Molecular Dynamics and Density Functional Theory.

2.1 Introduction to Molecular Dynamics

This section describes the fundamental principles of Molecular Dynamics (MD) simulations, focusing on basis assumptions and approximations.

2.1.1 Assumptions of Molecular Dynamics Simulations

Molecular Dynamics is a computational technique widely used in physics, chemistry, and biology to investigate the time evolution of complex molecular systems, ranging from condensed matter to biological macromolecules, within a particle-based description.

The overall effectiveness of molecular dynamics depends on the validity of its underlying principles. Four of them are reported in the following lines. The first assumption consists in the validity of the Born-Oppenheimer approximation of the Schrödinger Equation during a MD simulation of a system of particles. In principle, a fully relativistic quantum-mechanical description would provide the most rigorous representation of a molecular system. However, in such case the number of degrees of freedom would be extremely large, and thus its treatment becomes computationally intractable. Instead, the Born-Oppenheimer approximation assumes that first only electrons are considered as independent variables of $\mathcal{H}$, while positions of the nuclei are assumed fixed. This is justified by the different timescales characterizing the motion of electrons, that vibrate much faster than nuclei and can be even motivated by the discrepancy between the masses of a single electron and a nucleus ($m_{electron} \propto \frac{1}{2000} m_{proton}$). Moreover, at typical biochemical temperatures ($T \approx 300$ K), the thermal energy is $k_B T \approx 2.5 \text{ kJ mol}^{-1} \approx 0.6 \text{ kcal mol}^{-1} \approx 0.026 \text{ eV}$, which is several hundred times smaller than a typical electronic excitation energy (for example, $\Delta E \sim 10 \text{ eV}$ in the hydrogen atom). Thus, electronic dynamics is

solved independently according to QM, through a Schrödinger Equation. Its solution represents the electronic energy levels of a molecule as a function of atomic coordinates and is treated as an effective potential energy to investigate the motion of the nuclei. In the MD approach, the system is treated classically where electrons are implicit variables of an effective potential model (force field) and specific partial charges (representing the electronic potential) are fitted to the nuclei centers. Therefore, MD simulations benefit from previous fitted calculations deriving from quantum mechanics, retaining a classical approach [104].

The second fundamental principle is described by the thermodynamic hypothesis. It assumes that macromolecules are driven naturally to their native structure largely on the grounds of thermodynamics. This behavior is in contrast to biological processes, that often require the crucial interplay of enzyme machinery to lower activation barriers, in order to bring proteins to their functional state. Though it is important to consider these complex factors when modeling macromolecules, the basic thermodynamic hypothesis of the empirical approach appears reasonable for many biomolecules under study by molecular mechanics and dynamics [104].

The third assumption is the molecular mechanics principle of additivity. The effective molecular energy can be expressed as a sum of potentials derived from simple physical forces: Van der Waals, electrostatic, mechanical-like strains arising from "ideal" bond length and angle deviations, and internal torsion flexibility (rotation of two chemical groups around the bond joining them) [104].

Lastly, the principle of transferability assumes that potentials can be developed to incorporate all experimental data for representative structures and then be applied successfully to the prediction of large biological molecules composed of the same chemical subgroups. This is basically a reasonable assumption since bond lengths and bond angles tend to adopt similar values in different molecular species under normal conditions. However, under special straining forces, these values may vary significantly from 'ideal' values. For example, modeling the structures and properties of complex aromatic or conjugated systems requires special treatment such as using sophisticated quantum-mechanical calculations for parameterization [104].

2.1.2 The Born-Oppenheimer Approximation

According to the solution of the Born-Oppenheimer approximation scheme, the quantum description of the full system reduces to a classical treatment of the nuclei. The Hamiltonian of a molecular system can be written as

$$\mathcal{H}(\tilde{P}, \tilde{X}) = T(\tilde{P}) + U(\tilde{X}) \quad (2.1)$$

where $\tilde{X} = \{\tilde{X}_i\}_{i=1}^N$ and $\tilde{P} = \{\tilde{P}_i\}_{i=1}^N$ denote the nuclear coordinates and momenta of the N particles, respectively. The kinetic energy is

$$T(\tilde{P}) = \sum_{i=1}^N \frac{\tilde{P}_i^2}{2M_i} \quad (2.2)$$

with M_i the mass of the i -th nucleus. Within the Born–Oppenheimer approximation, the potential energy depends only on the nuclear coordinates and can be expressed as

$$U(\tilde{X}) = V_{nn}(\tilde{X}) + E_{\text{el}}^0(\tilde{X}) \quad (2.3)$$

where $V_{nn}(\tilde{X})$ represents the direct nucleus–nucleus interaction, and $E_{\text{el}}^0(\tilde{X})$ is the electronic ground-state energy, defining the potential energy surface. The classical motion of the nuclei follows Newton’s equations:

$$M_i \frac{d^2 \tilde{X}_i}{dt^2} = -\nabla_i \left[V_{nn}(\tilde{X}) + E_{\text{el}}^0(\tilde{X}) \right] \quad (2.4)$$

The electronic contribution to the potential energy surface is obtained by solving the time-independent Schrodinger equation at fixed nuclear coordinates:

$$\left[T_e(\mathbf{r}) + V_{ee}(\mathbf{r}) + V_{ne}(\mathbf{r}; \tilde{X}) \right] \psi_0(\mathbf{r}|\tilde{X}) = E_{\text{el}}^0(\tilde{X}) \psi_0(\mathbf{r}|\tilde{X}) \quad (2.5)$$

Here, $\mathbf{r}$ denotes the electronic coordinates, $\nabla^2 T_e$ the electronic kinetic energy operator, V_{ee} the electron–electron Coulomb interaction, and V_{ne} the electron–nucleus interaction. The nuclear positions enter as fixed parameters, meaning that electrons are assumed to instantaneously relax to their ground state whenever the nuclei move. This formalism defines Born–Oppenheimer Molecular Dynamics. Although conceptually rigorous within the approximation, its application to large systems is computationally demanding, since the electronic ground-state energy (and its derivatives) must be evaluated at every nuclear configuration along the trajectory.

2.2 The Molecular Dynamics Algorithm

The Molecular Dynamics algorithm generates a time evolution of a molecular system. The main elements required to perform a molecular dynamics simulation are the following:

1. atomic coordinates:
2. a force field, a set of both functions and parameters for energy calculation;

3. an integrator, an algorithm that iteratively updates positions and velocities of each particle, according to the potential energy defined in the force field.

2.2.1 Classical Force Fields

Force fields define a set of parameters that model the potential energy surface of a given system. These parameters are often derived from experimental data, which is why such force fields are referred to as “semi-empirical”. From physical theory, the derivative of the potential energy with respect to atomic positions corresponds to a negative force. Estimating these forces between system particles is essential for integrating their motions. A key assumption of standard force fields is that every potential energy function must have an analytic form that is continuously differentiable. Consequently, in MD simulations all forces governing the atomic behavior are conservative. This approximation contrasts with natural systems; for instance, solvents exhibit non-conservative effects such as viscosity and friction. Nevertheless, when treating systems at molecular resolution, these effects are often negligible, and the approximation remains reliable [105]. Each functional form defining a potential energy term in MD simulations is based on the principle of additivity. Accordingly, the effective molecular energy is expressed as a sum of potentials arising from simple physical interactions, both molecular and supramolecular, as shown in equation (2.6).

$$U_T = U_B + U_{NB} \quad (2.6)$$

where U_T is the total potential energy of the system, U_B , U_{NB} represent the interactions between covalently-linked and between non-covalently-linked atoms, respectively. These two energies can be further analysed, starting by U_B . The definition of intra-molecular potentials comprises the three principal bonded interactions U_{BL} , U_{BA} , U_{BT} , bond length, angle bending and dihedral angle torsion, respectively (equation (2.7)):

$$U_B = U_{BL} + U_{BA} + U_{BT} \quad (2.7)$$

Two atoms a, b covalently bound have an equilibrium bond length, $\bar{r}_{a,b}$. Such value minimizes the potential energy function describing their bond, and assuming the dissociation energy for the bond to be positive, the minimum is zero. To have the potential energy at an arbitrary point, it is possible to use a Taylor expansion about $\bar{r}_{a,b}$, truncating it after the first non-null term. The result is illustrated in equation (2.8) and defines the simplest possible expression for the vibrational potential energy [105].

$$U_{a,b} = k_{a,b} (r_{a,b} - \bar{r}_{a,b})^2 \quad (2.8)$$

where $r_{a,b}$ is the current distance between the two atoms, $\bar{r}_{a,b}$ the equilibrium bond length. In principle, $k_{a,b}$ should represent the numerical value of the second derivative of U_{ab} with respect position, evaluated in $\bar{r}_{a,b}$. However, it can be seen as the ‘force constant’ for the Hooke’s law for a spring [105]. A graphical image of U_{ab} is shown in Fig. 2.1.

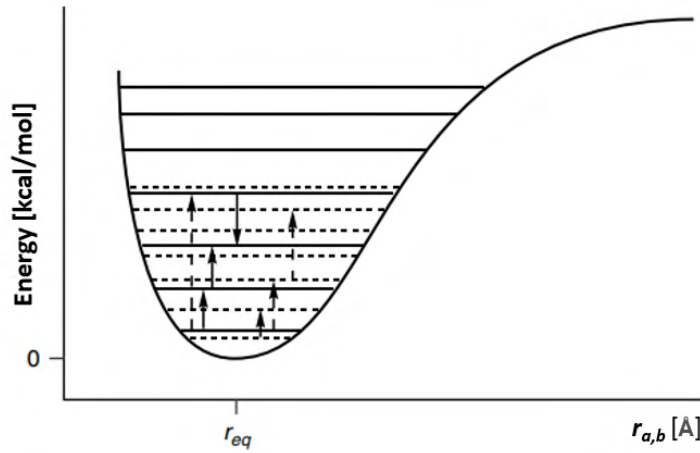


Figure 2.1: The first seven vibrational energy levels for a lighter (solid horizontal lines) and heavier (horizontal dashed lines) isotopomer of diatomic ab , where a , b are two different atoms. The curve in the figure indicates $U_{a,b}$, and r_{eq} , $r_{a,b}$ represent the equilibrium and current position, respectively, of the two atoms a , b . Allowed vibrational transitions are indicated by solid vertical arrows, forbidden transitions are indicated by dashed vertical arrows [105].

Therefore, the total bond length potential U_{BL} is given by the sum over all the pair terms $U_{a,b}$ of all possible bonds within the molecule:

$$U_{BL} = \sum_{a,b} k_{a,b} (r_{a,b} - \bar{r}_{a,b})^2 \quad (2.9)$$

However, equation (2.9) has some limitations. Like any truncated Taylor expansion, it works best in regions near the equilibrium point, $\bar{r}_{a,b}$. Thus, if interested primarily in molecular structures where no bond is terribly distorted from its optimal value, the equation has reasonable utility. However, as the bond gets decisively stretched, the predicted energy becomes infinitely positive, which is definitely not chemically realistic. One practical solution to such inaccuracy is to include additional terms in the Taylor expansion, but this inevitably leads to an unintended complication. There are other descriptions for U_{BL} , like the Morse potential, reported in equation (2.10), where $D_{a,b}$ is the dissociation energy of the bond and α_{ab} is a fitting constant, two

parameters estimated from experiments [105]:

$$U_{BL} = \sum_{a,b} D_{a,b} \{1 - \exp[-\alpha_{ab}(r_{a,b} - \bar{r}_{a,b})]\}^2 \quad (2.10)$$

The second term U_{BA} is associated, for small displacements from equilibrium, with energy variations caused by bond angle deformation. As well, it can be modeled by polynomial expansions as oscillations related to bond stretching, by truncating the Taylor expansion at a desirable level, e.g. first non-zero term to obtain a harmonic description (equation (2.11)), or by using trigonometric functions, like in equation (2.12).

$$U_{BA} = \sum_{a,b,c} k_{a,b,c} (\theta_{a,b,c} - \bar{\theta}_{a,b,c})^2 \quad (2.11)$$

$$U_{BA} = \sum_{a,b,c} k_{a,b,c} (\cos \theta_{a,b,c} - \cos \bar{\theta}_{a,b,c})^2 \quad (2.12)$$

where $\theta_{a,b,c}$ is the valence angle between bonds a - b , b - c and $\bar{\theta}_{a,b,c}$ the relative equilibrium angle. Moreover, $\cos \theta_{a,b,c}$ is defined as the dot product $(\mathbf{x}_a - \mathbf{x}_b) \cdot (\mathbf{x}_b - \mathbf{x}_c)$, normalized by the bond lengths $r_{a,b}$ and $r_{b,c}$, where $\mathbf{x}_i$ is the position vector of the i^{th} atom considered. In a force field, bonds are defined as vector connecting two atoms, so there is no ambiguity about what is meant by an angle between two bonds.. The force constants $k_{a,b,c}$ now directly depend on the three atoms a , b , c [105, 106].

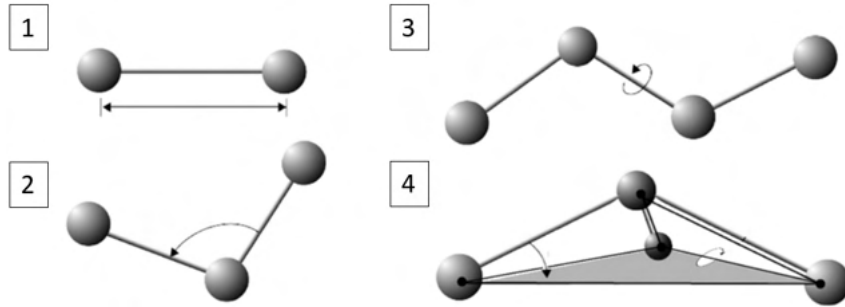


Figure 2.2: Schematic representation of bonded interactions. 1) bond stretch; 2) angle bending; 3) proper dihedral angle torsion; 4) improper dihedral angle torsion [105, 106].

Dihedral or torsional angles are illustrated in *figure 2.2 (3)*. The torsional potential U_{BT} refers to four atoms connected in sequence, a - b - c - d and the relative torsional angle is defined as the angle $\omega_{a,b,c,d}$ between bonds a - b , c - d , when they are projected into the plane bisecting the b - c bond. The convention is to define the angle as positive if one must rotate the bond in front of the bisecting plane in a clockwise fashion to eclipse the bond behind the bisecting plane. By

construction, the torsion angle $\omega_{a,b,c,d}$ is periodic. By convention, $-\pi \leq \omega_{a,b,c,d} \leq \pi$. Since the torsion itself is periodic, so too must be the torsional potential energy. Therefore, its function is modeled as an expansion of periodic functions, e.g., a Fourier series (equation (2.13)) [105, 106]:

$$U_{BT} = \frac{1}{2} \sum_{a,b,c,d} \sum_{j \in J} (V_{a,b,c,d}^j [1 + (-1)^{(j+1)} \cos(j \omega_{a,b,c,d} - \phi_{a,b,c,d}^j)]) \quad (2.13)$$

where the values of amplitudes (or height barriers) V_j and the set J included in the sum represent the periodicity of the torsional. Deleting a particular value of j from the evaluated set is equivalent to setting the term amplitude for that value of j equal to zero. The factor of $(-1)^{(j+1)}$ is included so that the function in brackets within the sum is zero for all j when $\omega_{a,b,c,d} = \pi$, if the phase angles $\phi_{a,b,c,d}^j$ are all set to 0. The dihedral angle $\omega_{a,b,c,d}$, can be expressed as reported in equation (2.14):

$$\omega_{a,b,c,d} = \arccos \left[\frac{\mathbf{r}_{ab} \times \mathbf{r}_{bc}}{\|\mathbf{r}_{ab}\| \|\mathbf{r}_{bc}\| \sin \theta_{\mathbf{r}_{ab}, \mathbf{r}_{bc}}} \bullet \frac{\mathbf{r}_{bc} \times \mathbf{r}_{cd}}{\|\mathbf{r}_{bc}\| \|\mathbf{r}_{cd}\| \sin \theta_{\mathbf{r}_{bc}, \mathbf{r}_{cd}}} \right] \quad (2.14)$$

While the mathematical utility of equation (2.13) is clear, it is also well founded in a chemical sense. In fact, U_{BT} considers the presence of torsional energy barriers, describing configurations that more energetically favored than others. A typical example is the functional expressing the butane dihedral angle potential, obtained by treating the parameters of the equation (2.13), as shown in Fig. 2.3. A separate situation, arises in the case of four-atom systems where a central atom is bonded to three otherwise unconnected atoms, e.g. formaldehyde. These particular dihedral configurations are called improper dihedrals or torsions, see Fig. 2.2 (4). Many force fields treat such improper torsions like any other torsion, i.e. using equation (2.13) [105]. Some other alternatives to model such cases exist, but will not be considered in this chapter.

The second term U_{NB} associated to non-bonded interactions in equation (2.6) can be split in two contributions, electrostatic and van der Waals respectively, as shown in the equation (2.15):

$$U_{NB} = U_{El} + U_{LJ} \quad (2.15)$$

The van der Waals term U_{LJ} models the Lennard-Jones potential. Because the motions of electrons are correlated (more properly, the electronic wave functions are correlated, according to QM), the two atoms a , b simultaneously develop electrical moments that are oriented so as to be mutually attractive. The force associated with this interaction is variously defined as ‘dispersion’, the ‘London’ force, or the ‘attractive van der Waals’ force. In the absence of a permanent charge, this interaction can be described as a dipole–dipole, usually referred to as an “induced dipole–induced dipole” interaction, since the moments in question are not permanent.

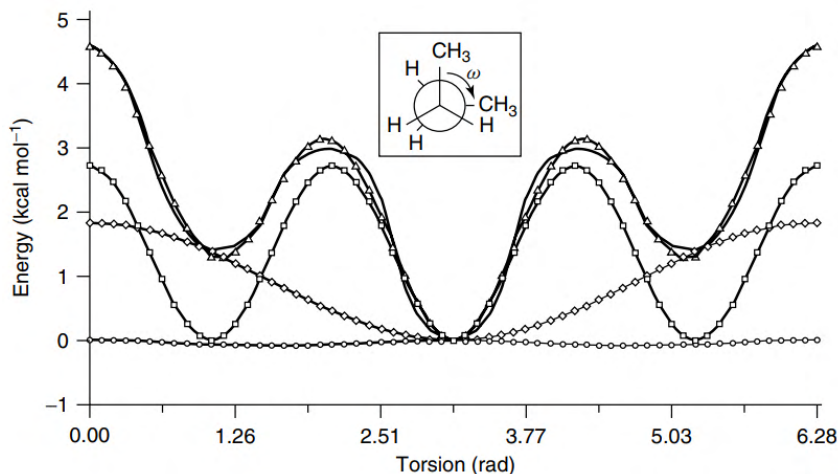


Figure 2.3: Fourier decomposition of the torsional energy for rotation about the C–C bond of n-butane (bold black curve, energetics approximate). The Fourier sum has a close overlap, and is composed of the onefold, twofold, and threefold periodic terms, respectively [105, 106].

Such an interaction has an inverse sixth power dependence on the distance between the two atoms a , b . Thus, the potential energy becomes increasingly negative as the two atoms approach one another from infinity. However, as the two atoms continue to approach one another, their surrounding electron densities ultimately begin to interpenetrate. In the absence of opportunities for bonding interactions, Pauli repulsion (or ‘exchange repulsion’) causes the energy of the system to rise rapidly with decreasing bond length (inverse twelve power dependence). Dispersion and repulsion respectively can be translated in a simple functional form for the U_{LJ} contribute, that sums up the two opposite effects for each atomic van der Waals pair r_{ab} present in the system, as reported in equation (2.16):

$$U_{LJ} = \sum_{a,b} \left(\frac{A_{ab}}{r_{ab}^{12}} - \frac{B_{ab}}{r_{ab}^6} \right) \quad (2.16)$$

where r_{ab} represents the distance between the two atom nuclei, and A_{ab} , B_{ab} are two constants, dependent on the nature of a , b , for repulsion and dispersion effects, respectively. In a formal analysis, the number of non-bonded interactions that must be evaluated scales as N^2 , where N is the number of atoms. In the process of optimizing geometry, or of searching for many energy minima for a complex molecule, hundreds or thousands of energy evaluations may need to be performed for interim structures. Thus, seemingly small savings in time can be done in code development, i.e. rewriting A_{ab} as $4\epsilon_{ab}\sigma_{ab}^{12}$ and B_{ab} as $4\epsilon_{ab}\sigma_{ab}^6$. This results in storing only two parameters per each atom couple, ϵ_{ab} and σ_{ab} .

Finally, the last term to consider in equation (2.15) is U_{El} , referred to the electrostatic

interactions. Classical electrostatics would tend to describe the related energy with multipole moments theory. However, this description is not very convenient in the context of intramolecular energy contributes, because such an approach would be computationally quite intensive, as evaluation of the moment potentials is tedious.

Therefore, the simplest approach to model electrostatic interactions is to assign to each atom a partial charge, and treat the relative energy between every pair of atoms a , b as Coulomb potentials, shown in equation (2.17).

$$V_{El} = \sum_{a,b} \left(\frac{q_a q_b}{\epsilon_{ab} r_{ab}} \right) \quad (2.17)$$

where r_{ab} is the distance between the two particles, ϵ_{ab} is the permittivity, a measure of the encountered resistance if generating an electric field in a given medium (proportional to the relative dielectric constant), q_a and q_b are the full or partial charges of the atom a and b respectively, and the sum accounts for all the contributes of atoms within the ensemble related to supramolecular interactions. The charge of an atom is permanent, so once defined, that atom must carry that charge along the whole simulation. Thus, the atomic charge is a fixed parameter [105, 106].

A major difficulty in the computation of electrostatic interactions arises from the long-range nature of the Coulomb potential, which decays as $1/r$ and therefore cannot be safely truncated at a finite cutoff distance. This issue becomes particularly critical in simulations performed under periodic boundary conditions, where each charge interacts not only with particles in the primary simulation box but also with their infinite periodic images.

A rigorous and widely adopted strategy to address this problem is the Ewald summation method. The central idea consists in adding and subtracting, at each point charge, a Gaussian charge distribution of equal magnitude and opposite sign. Since the two Gaussian distributions cancel each other, the total charge density of the system remains unchanged. However, this mathematical manipulation allows the electrostatic energy to be decomposed into contributions that are more convenient to evaluate. The total electrostatic energy can thus be written as:

$$U_{ele}(\{\mathbf{r}_i\}) = U_{real}(\{\mathbf{r}_i\}) + U_{rec}(\{\mathbf{r}_i\}) + U_{self} \quad (2.18)$$

The first term, U_{real} , represents the interaction between point charges screened by Gaussian distributions of opposite sign. Because these screened charge distributions are overall neutral and possess only multipole moments, their interactions decay faster than $1/r$ and can therefore be computed efficiently in real space using a finite cutoff. The second term, U_{rec} , accounts for the interaction between the compensating Gaussian charge distributions. Owing to their smooth and spatially extended nature, this contribution is most efficiently evaluated in reciprocal (Fourier)

space, where periodic boundary conditions can be naturally exploited. The third term, U_{self} , is a self-interaction correction that subtracts the artificial interaction between each point charge and its own compensating Gaussian distribution. This term does not depend on the particle positions and is computed only once.

An efficient implementation of the Ewald approach is the Particle Mesh Ewald (PME) method. In PME, charges are interpolated onto a regular three-dimensional grid, and the reciprocal-space contribution is computed using Fast Fourier Transform (FFT) algorithms, resulting in a computational scaling of order $\mathcal{O}(N \log N)$. This method is implemented in GRO-MACS, the molecular dynamics software package used to generate the simulation data presented in this thesis.

Lastly, several additional aspects require consideration. Bonds, angles, and torsions are not independent molecular coordinates, as they are coupled with one another. Cross terms, which depend on variables of different types (e.g., distance and angle between two atoms), are often included in force fields to more accurately model specific biologically relevant interactions. This extends equation (2.6), adding further terms to the total potential energy of the system, U_T . Moreover, additional non-bonded terms, such as those specifically modeling different types of dipoles, can be incorporated. The charge distribution around each nucleus, as described in equation (2.17), is generally complex, and in some cases, induced dipole effects must be accounted for. It is important to note that a force field is a large collection of functional forms and associated parameters. With this collection, the energy of a given molecule—whose atomic connectivity is typically specified—can be evaluated by computing the energy contributions of all defined interactions present in the molecule. Although the number of such interactions can be substantial, the computations are straightforward and easily handled by a digital computer. However, the individual parameters in the functional forms (e.g., force constants, equilibrium coordinate values, phase angles) are not always directly measurable by spectroscopy. The parameterization process is therefore crucial in the construction of force fields, as previously introduced in the discussion of van der Waals interactions. It becomes clear that atomic “types” cannot be determined by atomic number alone. For example, bonds between two carbon atoms can belong to at least four distinct classes—single, aromatic, double, and triple—each with its own stretching force constant and equilibrium distance. A similar situation arises for any atom pair when multiple bonding is possible, and different atomic hybridizations lead to distinct angle-bending equilibrium values [105, 106]. This creates an underdetermined optimization problem when finalizing a force field. One possible approach is to make certain atom-pair parameters explicitly dependent on the properties of the individual atoms, effectively rendering them functions of single-atom-specific parameters. For instance, the two van der Waals parameters σ_{ab} and ϵ_{ab}

can be expressed as shown in equations (2.19) and (2.20):

$$\sigma_{ab} = \sigma_a + \sigma_b \quad (2.19)$$

$$\epsilon_{ab} = (\epsilon_a \epsilon_b)^{\frac{1}{2}} \quad (2.20)$$

where σ_a , ϵ_a are specific constants for atom a (the same for constants referred to b). This reduces the need for $\frac{N(N+1)}{2}$ diatomic parameters to only N atomic parameters, in each atom pair case, where N is the number of different atom types within the system considered (e.g. C- α , C_{carboxile}, etc...). However, these parameters often vary according to the force field model. In fact, no universal force-field exists, and different force-field variants are available, all describing different parametrization according to their own rules. Given a system, its compatibility with a certain potential model must be tested. Generally, the three types of force fields currently are developed: All Atom, in which parameters are provided for every single atom within the system; United Atom, in which parameters are provided for all atoms except non-polar hydrogens; Coarse Grained models, characterized by an abstract representation of molecules by clustering several atoms into bigger beads, that constitute rigid architectures, enhancing the performance of the simulation but reducing the system resolution in terms of chemical description and accuracy.

2.2.2 Reactive Force Fields

Force fields can also be used to model chemical reaction, in a complete different framework, called Reactive Molecular Dynamics. In this section, a specific type of the reactive force field will be introduced, namely ReaxFF [107]. ReaxFF is an atomistic simulation framework designed to model chemical reactivity within a classical Molecular Dynamics (MD) scheme. Unlike classical force fields, which rely on predefined bond connectivity, ReaxFF employs a bond-order formalism in which bonds are not fixed but continuously evolve as a function of interatomic distances. This allows for the seamless description of bond formation and breaking events during a simulation [108].

From a theoretical standpoint, ReaxFF bridges the gap between quantum mechanical (QM) methods and classical force fields. While QM approaches explicitly treat electrons and provide high accuracy, they are computationally prohibitive for large systems and long timescales. ReaxFF, instead, encodes chemical information into an empirical potential trained on QM data, enabling simulations of reactive processes at significantly reduced computational cost [108].

The accuracy of ReaxFF strongly depends on the quality and transferability of the training dataset used during parametrization. Since the potential is fitted to reproduce specific QM observables (e.g., energies, geometries, reaction barriers), it may lose predictive power outside its training domain. This introduces intrinsic uncertainties, particularly in chemically hetero-

geneous systems. Additional limitations arise from the charge equilibration scheme (typically based on electronegativity equalization), which may lead to unphysical long-range charge transfer in some cases [108]. Consequently, while ReaxFF provides a powerful framework for reactive simulations, but its predictions must always be interpreted with care.

Unlike classical force fields, where the connectivity between atoms is fixed, ReaxFF allows for the continuous formation and breaking of chemical bonds during the simulation. At the core of ReaxFF lies the concept of a distance-dependent bond order. For a pair of atoms i and j , the total bond order BO_{ij} is expressed as the sum of σ , π , and double- π contributions:

$$BO_{ij} = BO_{ij}^{\sigma} + BO_{ij}^{\pi} + BO_{ij}^{\pi\pi} \quad (2.21)$$

Each contribution is modeled as an exponential function of the interatomic distance r_{ij} :

$$BO_{ij}^{\sigma} = \exp \left[p_{bo1} \left(\frac{r_{ij}}{r_0^{\sigma}} \right)^{p_{bo2}} \right] \quad (2.22)$$

$$BO_{ij}^{\pi} = \exp \left[p_{bo3} \left(\frac{r_{ij}}{r_0^{\pi}} \right)^{p_{bo4}} \right] \quad (2.23)$$

$$BO_{ij}^{\pi\pi} = \exp \left[p_{bo5} \left(\frac{r_{ij}}{r_0^{\pi\pi}} \right)^{p_{bo6}} \right] \quad (2.24)$$

where every p_{boK} , with $K = 1, 2, 3, 4, 5, 6$, is a fitting parameter reproducing bond stiffness. These expressions ensure that the bond order smoothly decreases with increasing interatomic distance, enabling a continuous transition between bonded and non-bonded states. The total energy of the system is written as a sum of physically meaningful contributions:

$$E_{\text{tot}} = E_{\text{bond}} + E_{\text{over}} + E_{\text{under}} + E_{\text{val}} + E_{\text{tors}} + E_{\text{vdW}} + E_{\text{Coulomb}} \quad (2.25)$$

The bond energy term depends explicitly on the bond order:

$$E_{\text{bond}} = \sum_{i < j} D_e^{ij} \cdot BO_{ij} \cdot \exp[-p_{be1}(1 - BO_{ij})] \quad (2.26)$$

Here, D_e^{ij} represents the bond dissociation energy, while p_{be1} is a fitting parameter controlling bond stiffness. To ensure chemically meaningful coordination, ReaxFF includes penalty terms for over- and under-coordination:

$$E_{\text{over}} = \sum_i p_{ov}(CN_i - CN_i^{\text{opt}})^2 \quad (2.27)$$

$$E_{\text{under}} = \sum_i p_{\text{un}}(CN_i^{\text{opt}} - CN_i)^2 \quad (2.28)$$

where $CN_i = \sum_j BO_{ij}$ is the coordination number of atom i . Angular and torsional interactions are also bond-order dependent:

$$E_{\text{val}} = \sum_{i,j,k} f(BO_{ij}, BO_{jk}) \cdot k_{\theta}(\theta_{ijk} - \theta_0)^2 \quad (2.29)$$

$$E_{\text{tors}} = \sum_{i,j,k,l} f(BO_{ij}, BO_{jk}, BO_{kl}) \cdot V_n [1 + \cos(n\phi - \gamma)] \quad (2.30)$$

Non-bonded interactions are treated through van der Waals and Coulomb terms:

$$E_{\text{vdW}} = \sum_{i<j} D_{ij} [\exp(-\alpha r_{ij}) - 2 \exp(-\beta r_{ij})] \quad (2.31)$$

$$E_{\text{Coulomb}} = \sum_{i<j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.32)$$

The partial charges q_i are dynamically updated at each timestep using a charge equilibration (QEq) scheme:

$$E_{\text{charge}} = \sum_i \left(\chi_i q_i + \frac{1}{2} J_i q_i^2 \right) + \sum_{i<j} \frac{q_i q_j}{R_{ij}} \quad (2.33)$$

where χ_i is the electronegativity and J_i the hardness of atom i .

A defining feature of ReaxFF is the use of a *single atom type per element*, independent of the local chemical environment. For instance, all oxygen atoms are treated identically whether they belong to water, silica, or a carbonyl group. This contrasts with classical biomolecular force fields, where multiple atom types are defined for the same element depending on hybridization or bonding context. While this choice enables chemical flexibility and phase transitions, it also introduces limitations in accuracy, as subtle environmental effects are averaged into a single parametrization.

2.2.3 Integrators

Moving atoms within a system is a difficult task. In principle, calculating the partition function and associated thermodynamic properties for a general many-body potential that includes non-linear interactions becomes an insurmountable task if only analytical techniques are employed. Unless a system can be transformed into a more tractable form, it is very unlikely that this task can be performed analytically. A possible solution could be introducing approximations, exploiting a simpler model system, or employing numerical methods, like an integrator [104].

Integrators offer the possibility to update the positions and velocities of the atoms within a system, provided a description of the potential energy, the force field. One principal characteristic of an integrator is time reversibility: a backwards integration of a forward-integrated trajectory passes through the same points, but in the reverse order, given both reverse and forward trajectories are computed with the same method. Moreover, energy conservation principles should be respected, according to the simulated ensemble. For example, the total energy variation of a system is null during time in an NVE ensemble. Contrary, in settings like NPT or NVT, its instant value is allowed to change, but the relative fluctuations are so tiny in respect to the average value, that these deviations are considered negligible, given that they approximately scale with $\frac{1}{\sqrt{N}}$, where N is the total number of particles within a system. Furthermore, given an infinite amount of time, a system with a given energy is able to visit all configurations on the constant energy hypersurface. A system with this property is said to be ergodic and can be used to generate a microcanonical ensemble (NVE). Ergodicity is always assumed when performing MD and generating the relative trajectories. An integrator generates phase space vectors at discrete times that are multiples of a fundamental time discretization parameter, Δt , known as the time step. Starting with the initial condition $\mathbf{x}_0$, phase space vectors $\mathbf{x}_{n\Delta t}$, where $n = 0, \dots, M$ are generated by applying the integrator or solver iteratively [104]. By far the simplest way to obtain a numerical integration scheme is to use a Taylor series. A common one is Velocity Verlet. The coordinate file in input specifies the initial position vector $\mathbf{x}_0$, while initial velocities of a particle i are instead generated by sampling the Boltzmann distribution at a given reference temperature, as follows in the equation (2.34).

$$p(v_i) = \sqrt{\frac{m_i}{2\pi k_b T}} \exp\left(\frac{-m_i v_i^2}{2k_b T}\right) \quad (2.34)$$

Given the initial positions and velocities $x(t=0)$ and $v(t=0)$, the Velocity-Verlet algorithm (a widely used integrator) integrates numerically the Newton's Equation as follows, generating a trajectory of the particles (described here below).

Velocity Verlet Algorithm

Initialization

$$\mathbf{x}(0), \quad \mathbf{v}(0)$$

Compute initial acceleration

$$\mathbf{a}(0) = \frac{1}{m} \nabla V(\mathbf{x}(0))$$

Step 1: Half-step velocity

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}(t) + \frac{\Delta t}{2}\mathbf{a}(t)$$

Step 2: Position update

$$\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right) \Delta t$$

Step 3: New acceleration

$$\mathbf{a}(t + \Delta t) = \frac{1}{m} \nabla V(\mathbf{x}(t + \Delta t))$$

Step 4: Final velocity

$$\mathbf{v}(t + \Delta t) = \mathbf{v}\left(t + \frac{\Delta t}{2}\right) + \frac{\Delta t}{2}\mathbf{a}(t + \Delta t)$$

Loop

$$t \leftarrow t + \Delta t$$

Repeat the procedure.

The typical Δt used is around 1-2 [fs]. The number of iterations required to achieve biologically relevant timescales is therefore huge: 1000 steps are needed to reach 1 [ns] of dynamics with $\Delta t = 1$ [fs]. Given a set of velocities at a certain time t , this algorithm updates the particle position vector, that defines together with the momenta a state of the system at that time instance. Then, forces are derived from the updated positions and the gradient of the potential energy computed from the force field, as in equation (2.35).

$$\mathbf{F}(\mathbf{x}(t)) = -\nabla U(\mathbf{x}(t)) \quad (2.35)$$

The updated forces are then used to compute new velocities, referred to the end of this particular iteration cycle [104].

2.2.4 Periodic Boundary Conditions and Simulation Box

The atoms of the simulated system are placed into a space-filling box, which is surrounded by translated copies of itself, also called periodic boundaries (PBC) that usually span over all the three dimensions. For example, in order to simulate in explicit solvent, the protein is inserted in a box surrounded by water molecules; also ions are often present to balance residual sources of charge (a system must be neutral), or to emulate physiological cell concentrations. The three most useful box types for simulations of solvated systems are: cubic, rhombic dodecahedron and truncated octahedron [104]. In this study, cubic boxes are utilized. An image is shown in Fig. 2.4.

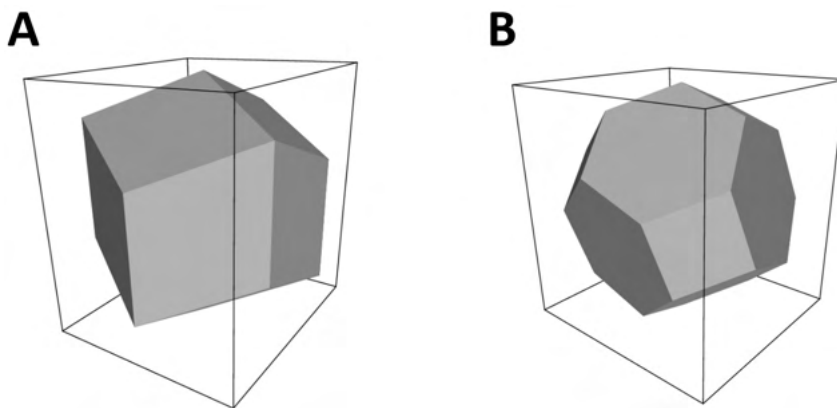


Figure 2.4: Typical box shapes utilized in Molecular Dynamics simulations: rhombic dodecahedron (A) and truncated octahedron (B) simulation boxes [104].

2.2.5 Thermostats and Barostats

The numerical integration of Newton's equations of motion treated in the previous paragraphs is considered for an isolated system. Therefore, Newton's equations alone are insufficient for the accurate treatment of biomolecules. In fact, Molecular Dynamics has the objective to simulate biological systems *in silico*, and most of the conditions found on biological reactions involve a constant pressure or temperature. This means that the system can exchange energy with the surrounding environment. To account for these effects, thermodynamic constraints can be introduced, typically by controlling the temperature T and pressure P . Under these conditions, the sampled configurations $\{\mathbf{r}_i\}$ follow the isothermal–isobaric ensemble (NPT) distribution:

$$p(\{\mathbf{r}_i\}) = \frac{1}{Z} \exp \left[-\frac{U(\{\mathbf{r}_i\}) + PV(\{\mathbf{r}_i\})}{k_B T} \right], \quad (2.36)$$

where $U(\{\mathbf{r}_i\})$ is the potential energy, $V(\{\mathbf{r}_i\})$ is the system volume, k_B is the Boltzmann constant, and Z is the partition function. Algorithms have been developed to maintain the system at constant temperature and pressure, simulating the presence of a thermostat and a barostat in contact with the system. First, the relation between thermostat and temperature will be explained, and second the link between pressure and barostats.

Thermostats

Given the bound between temperature and the average kinetic energy explicated by the Virial theorem, thermostats act on the system by retaining this mean value of the kinetic energy around a constant input value. In equation (2.37) is reported such relation:

$$\left\langle \sum_{i=1}^{3N} \frac{\mathbf{p}_i^2}{2m_i} \right\rangle = \frac{3}{2} N k T \quad (2.37)$$

where $\mathbf{p}_i^2$ represents the squared i^{th} coordinate of the vector of momenta $\mathbf{p}$ (stores three entries (x, y, z) per particle), N is the particle number, k is the Boltzmann constant and T the temperature. The left-hand side represents the (empirical) average of the kinetic energy. In fact, kinetic energy directly depends on velocities. The simplest possible idea for a thermostat is represented by the Velocity-Rescaling, that applies, as the name suggests, velocity rescaling techniques in order to avoid that velocity (kinetic energy obtained from it) exceeds a certain value. The scaling factor α can be computed, as follows in equation (2.38).

$$\alpha = \sqrt{\frac{\tilde{K}}{K}} \quad (2.38)$$

where $\tilde{K}$ represents the target kinetic energy, computed from the input temperature value, and K the instant or empirical kinetic energy. Therefore, all atoms, whose velocities do not permit a constant value for the temperature, are rescaled. By this way, the temperature of the system is maintained constant. There are more complex thermostats, such as the Nosé-Hoover-chain or the Bussi Velocity Verlet scheme. A full treatment of these algorithms is out of the purpose of this manuscript, but a brief and short description will be given in the following paragraph.

A slightly more refined approach with respect to simple velocity rescaling is provided by the Berendsen thermostat, which introduces a weak coupling between the system and an external heat bath. Instead of enforcing an instantaneous rescaling of velocities, the Berendsen scheme relaxes the system temperature $T(t)$ towards a target temperature T_0 with a characteristic time

constant τ . This is achieved by modifying the scaling factor as:

$$\alpha = \sqrt{1 + \frac{\Delta t}{\tau} \left(\frac{T_0}{T(t)} - 1 \right)}, \quad (2.39)$$

where Δt is the simulation time step. In this way, temperature fluctuations are smoothly damped, avoiding the abrupt changes typical of simple velocity rescaling. However, it is important to note that the Berendsen thermostat does not generate a correct canonical ensemble, as it artificially suppresses natural fluctuations of the kinetic energy.

To overcome these limitations, stochastic approaches such as the Bussi velocity Verlet scheme (also known as stochastic velocity rescaling) have been developed. In this method, the velocity rescaling factor is modified by introducing a stochastic term that ensures the correct sampling of the canonical (NVT) ensemble. The evolution of the kinetic energy is governed by a stochastic differential equation, leading to a scaling factor α that depends not only on the instantaneous kinetic energy but also on a random variable drawn from an appropriate distribution. In the scheme proposed by Bussi *et al.* in Ref. [109], the kinetic energy is evolved so as to correctly sample the canonical ensemble by introducing a stochastic contribution to the rescaling factor. In particular, the velocities are updated according to

$$\mathbf{v}'_i = \alpha \mathbf{v}_i, \quad (2.40)$$

where the stochastic scaling factor α is given by

$$\alpha^2 = e^{-\Delta t/\tau} + \left(1 - e^{-\Delta t/\tau}\right) \frac{\tilde{K}}{K} + \frac{2e^{-\Delta t/\tau}}{N_f} \frac{\tilde{K}}{K} R + \frac{1 - e^{-2\Delta t/\tau}}{N_f} \frac{\tilde{K}}{K} R^2. \quad (2.41)$$

Here:

- K is the instantaneous kinetic energy,
- $\tilde{K} = \frac{N_f}{2} kT$ is the target kinetic energy,
- N_f is the number of degrees of freedom,
- τ is the thermostat relaxation time,
- Δt is the integration time step,
- R is a random number drawn from a standard normal distribution.

This stochastic term ensures that the fluctuations of the kinetic energy follow the correct canonical distribution, in contrast to deterministic rescaling schemes such as Berendsen, while maintaining a simple and efficient implementation. As a result, the Bussi thermostat preserves the

simplicity of velocity rescaling while correctly reproducing the Boltzmann distribution of kinetic energies. Furthermore, when combined with the Velocity Verlet integrator, it provides a numerically stable and efficient scheme widely used in molecular dynamics simulations, ensuring both proper thermodynamic sampling and good energy conservation properties.

Barostats

The derivation of NPT ensembles requires that volume is allowed to vary in order to keep the internal pressure equal, on average, to the applied external pressure. This suggests that to develop a molecular dynamics technique for generating isobaric ensembles, the volume must be introduced as an independent dynamical variable in the phase space. Indeed, the work-virial theorem, reported in equation (2.42) strongly supports such notion, since it effectively assigns an energy of kT to a "volume mode" [104].

$$\langle P^{(int)} V \rangle + kT = P \langle V \rangle \quad (2.42)$$

The idea of incorporating the volume into the phase space as an additional dynamical degree of freedom, together with its conjugate momentum, as a means of generating an isobaric ensemble was first introduced by Andersen (1980) and later generalized for anisotropic volume fluctuations by Parrinello and Rahman (1980) [104]. These two models were translated into barostats, that correspond most closely to laboratory conditions with a flask open to the external pressure. Practically, into the MD algorithm barostats are a set of equations that modify the equations of motion in order to allow the simulation box to change its shape to adapt the volume to a constant pressure. There are many types of barostats, even inspired by numerous other powerful techniques, based on extended phase spaces, including the Car-Parrinello approach (Car and Parrinello, 1985) for performing molecular dynamics with forces obtained from electronic structure calculations, and schemes for including nuclear quantum effects in molecular dynamics.

2.3 Coarse-Grained Molecular Dynamics

Up to now, Molecular Dynamics was described within the atomistic framework: every particle of the system resembles an atom of the modeled molecule, that is assigned to a specific parametrization strategy dictated by the chosen force field. Such framework is also known as Atomistic Molecular Dynamics. However, there are other ways to exploit statistical mechanics to produce reliable results and insights on biomolecules: among these, Coarse-Grained (CG) Molecular Dynamics. In the following sections, a brief and quick overview on CG MD will be given, mainly focusing on biological applications. Such section introduces the theoretical

principles described in Chapter 3, Section 3.6.2.

2.3.1 Motivation for Coarse-Grained Models

Atomistic MD provides a detailed and physically grounded description of molecular systems by explicitly representing every atom. However, this level of resolution comes with significant computational cost. The accessible time scales are typically limited to microseconds (in favorable cases), and system sizes rarely exceed millions of atoms without substantial computational resources. Many phenomena of biological and soft-matter relevance—such as membrane remodeling, protein aggregation, vesicle formation, and polymer self-assembly—occur on length scales of tens to hundreds of nanometers and time scales extending to milliseconds or longer. These regimes remain largely inaccessible to fully atomistic simulations.

Coarse-grained (CG) molecular dynamics addresses this limitation by reducing the number of degrees of freedom in the system. Groups of atoms are represented as single interaction sites (beads), thereby smoothing the energy landscape and allowing for larger integration time steps. This reduction enables simulations of larger systems and longer time scales, at the cost of reduced chemical detail. In other words, the access to larger length and time scales is guaranteed by a smoother energy landscape associated to a system displays less complex features, enabling faster sampling.

2.3.2 Principles of Coarse-Graining

The central idea of coarse-graining is to construct a mapping from an atomistic representation to a reduced model. Let $\mathbf{r}$ denote the full atomistic coordinates and $\mathbf{R}$ the coarse-grained coordinates. A mapping operator $\mathcal{M}$ defines the transformation:

$$\mathbf{R} = \mathcal{M}(\mathbf{r}). \quad (2.43)$$

Each coarse-grained bead typically represents a group of atoms, such as a methyl group, an amino acid side chain, or several water molecules. The choice of mapping is not unique and depends on the system and the phenomena of interest. Once the mapping is defined, an effective potential $U_{\text{CG}}(\mathbf{R})$ must be constructed to reproduce relevant properties of the underlying atomistic system. In principle, this potential corresponds to a potential of mean force:

$$U_{\text{CG}}(\mathbf{R}) = -k_B T \ln P(\mathbf{R}), \quad (2.44)$$

where $P(\mathbf{R})$ is the probability distribution of coarse-grained configurations obtained by projecting the atomistic ensemble. Formally, this exact coarse-grained potential is also written

as:

$$U_{\text{PMF}}(\mathbf{R}) = -k_B T \ln \int d\mathbf{r} \delta(\mathcal{M}(\mathbf{r}) - \mathbf{R}) e^{-\beta U(\mathbf{r})}, \quad (2.45)$$

where $U(\mathbf{r})$ is the atomistic potential energy and $\mathcal{M}$ is the mapping operator. This expression highlights two important aspects: the exact CG potential is generally *many-body* in nature, and it depends explicitly on thermodynamic conditions (e.g., temperature, pressure). In practice, CG models approximate U_{PMF} using simplified functional forms, typically limited to pairwise and low-order bonded interactions. The parameterization process thus seeks the best compromise between accuracy and computational efficiency. As explained in the case of atomistic MD, the effective potential can be typically decomposed into bonded and non-bonded contributions:

$$U_{\text{CG}} = U_{\text{bonded}} + U_{\text{non-bonded}}. \quad (2.46)$$

2.3.3 Dynamics in Coarse-Grained Models

The reduction of degrees of freedom alters the dynamical properties of the system. Because fast degrees of freedom are averaged out, CG models typically exhibit accelerated dynamics. As a result, time in CG simulations does not directly correspond to real physical time. To address this, time rescaling factors are often applied to the simulation time; another solution can be to use Langevin dynamics with adjusted friction coefficients, that allow to recover the putative real dynamics of the system. Nonetheless, interpreting dynamical observables requires caution, and is still a difficult task to address properly.

2.3.4 Parameterization Strategies in Coarse-Grained Models

The construction of a reliable coarse-grained (CG) force field critically depends on the parameterization procedure used to define the effective interactions between beads. Unlike atomistic force fields, where parameters are often associated with well-defined physical interactions (e.g., bonded terms, electrostatics), CG parameters are *effective* quantities that implicitly incorporate the influence of the degrees of freedom that have been integrated out. As a consequence, parameterization is inherently non-unique and must be guided by the specific properties one aims to reproduce. The main schemes with some examples will be rapidly introduced in the following lines. However, it is worth mentioning that no detailed description of CG will be given in this thesis.

Bottom-Up Approaches

Bottom-up methods derive CG potentials directly from atomistic simulation data, ensuring consistency with a reference high-resolution model. Lots of techniques can be utilized, here

there is a very quick and rapid overview, describing the most known and utilized ones.

Iterative Boltzmann Inversion (IBI) is one of the most widely used structure-based techniques. The goal is to reproduce target radial distribution functions $g_{\text{target}}(r)$ obtained from atomistic simulations. Starting from an initial guess (often the potential of mean force),

$$U_0(r) = -k_B T \ln g_{\text{target}}(r), \quad (2.47)$$

the potential is iteratively refined:

$$U_{n+1}(r) = U_n(r) + k_B T \ln \frac{g_n(r)}{g_{\text{target}}(r)}. \quad (2.48)$$

While IBI is conceptually simple and effective for simple liquids, it has limitations: it assumes pairwise additivity, may not reproduce thermodynamic properties accurately and convergence can be slow or unstable for complex large systems.

Another bottom-up parametrization scheme is the Force Matching (Multiscale Coarse-Graining). Force matching, introduced by Ercolessi and Adams [110], aims to reproduce the forces acting on CG sites as projected from atomistic simulations. The CG force field parameters are obtained by minimizing the objective function:

$$\chi^2 = \left\langle \sum_i |\mathbf{F}_i^{\text{atom}} - \mathbf{F}_i^{\text{CG}}|^2 \right\rangle, \quad (2.49)$$

where $\mathbf{F}_i^{\text{atom}}$ are the mapped atomistic forces. This method naturally incorporates many-body effects and is not limited to pairwise potentials. However it requires large amounts of atomistic trajectory data, and the resulting potentials may not reproduce structural distributions perfectly.

Lastly, also models based on entropy measurements were adopted. In fact, relative entropy approaches define the optimal CG model as the one minimizing the Kullback–Leibler divergence between atomistic and CG probability distributions:

$$S_{\text{rel}} = \int d\mathbf{R} P_{\text{atom}}(\mathbf{R}) \ln \frac{P_{\text{atom}}(\mathbf{R})}{P_{\text{CG}}(\mathbf{R})}. \quad (2.50)$$

This framework provides a unifying theoretical basis connecting structure-based and force-based methods [111]. In practice, it leads to an optimization problem over the CG parameters.

Top-Down Approaches

Top-down parameterization relies on matching experimental observables rather than atomistic simulations. These observables typically include:

- Densities and compressibilities
- Partition coefficients between phases
- Surface tensions
- Free energies of solvation

The advantage of top-down approaches is their direct connection to experimentally measurable quantities. However:

- They may lack structural fidelity at the microscopic level.
- Transferability can still be limited outside the calibration conditions.

Hybrid Approaches

Many modern CG force fields adopt a hybrid strategy, combining bottom-up and top-down elements. For example, bonded interactions may be derived from atomistic simulations, while non-bonded parameters are tuned to reproduce thermodynamic data. This combination allows to retain local structural accuracy while ensuring correct large-scale behavior.

Representability and Transferability

Two fundamental challenges arise in CG parameterization, involving both transferability and representability. In fact, a CG model with simplified interactions may not be able to simultaneously reproduce all properties of the atomistic system. For instance, a pairwise potential might match radial distribution functions but fail to reproduce pressure. Moreover, because CG potentials depend on thermodynamic conditions and chemical environment, parameters optimized for one state point may not be valid under different conditions (e.g., different temperatures or concentrations). These limitations are intrinsic to the coarse-graining process and must be carefully considered when designing and applying CG models. For example, it is well known that CG has difficulties in representing specific interactions like hydrogen bonding or electrostatics. Thus, CG models should be chosen carefully depending on the scientific question, and sometimes they cannot be used as reliable tools.

2.3.5 The MARTINI Force Field

The MARTINI force field is one of the most widely used coarse-grained models for biomolecular simulations. It was originally developed for lipid systems and later extended to proteins, carbohydrates, and other biomolecules [112]. MARTINI follows a top-down approach, where

interaction parameters are tuned to reproduce thermodynamic properties such as partitioning free energies between polar and apolar phases. This makes the force field particularly suitable for studying self-assembly and phase behavior.

Parametrization and Mapping Scheme

MARTINI parameterization is based on a fragment-based strategy in which small chemical moieties (building blocks) are mapped onto individual coarse-grained beads. Instead of attempting to reproduce detailed structural correlations, each bead type is calibrated to match a key thermodynamic observable: the free energy of transfer between different solvents, most commonly between water and an apolar phase such as octanol. In practice, this is achieved by tuning the Lennard-Jones interaction strengths (ϵ) so that the partitioning behavior of the coarse-grained bead mimics that of the corresponding chemical fragment. This choice is physically motivated, as partition coefficients encode the balance between hydrophilic and hydrophobic interactions, which are dominant drivers of self-assembly in soft-matter and biomolecular systems. By ensuring that each bead reproduces the correct solvation thermodynamics, the MARTINI model achieves a consistent and transferable description of intermolecular interactions across a wide range of environments, enabling realistic modeling of processes such as membrane formation, protein insertion, and phase separation.

MARTINI employs a relatively high level of coarse-graining, typically mapping four heavy atoms to one interaction bead. This 4:1 mapping strikes a balance between computational efficiency and chemical specificity. Beads are categorized into different types based on their polarity and chemical nature: Polar (P), Nonpolar (N), Apolar (C), Charged (Q). Each category is further subdivided to capture subtle differences in interaction behavior.

Non-Bonded Interactions

Non-bonded interactions are defined through a discrete interaction matrix between bead types. The Lennard-Jones well depth ϵ is varied systematically to represent different interaction strengths, while the bead size σ is typically fixed within classes. The utilized Lennard-Jones (LJ) potential is the same of equation 2.16, where in this case ϵ determines the interaction strength and σ the bead size.

Electrostatic interactions are described with a Coulomb potential, but with an effective relative dielectric constant typically set to $\epsilon_r = 15$, significantly larger than the vacuum value. This choice reflects a key aspect of CG: the explicit degrees of freedom responsible for dielectric screening—such as electronic polarization and the orientational response of atomistic water—are no longer resolved. Instead, their collective effect is incorporated implicitly into the model.

In atomistic simulations, the dielectric response of the environment emerges naturally from the explicit motion and polarization of molecules. In contrast, MARTINI uses a reduced representation in which each bead corresponds to several atoms, and water itself is modeled with limited polarity. As a consequence, electrostatic interactions would be severely overestimated if a low dielectric constant were used. The elevated value of ϵ_r effectively weakens Coulomb interactions, mimicking the screening that would otherwise arise from the neglected microscopic degrees of freedom. This treatment is consistent with the overall top-down philosophy of MARTINI, where interaction parameters are chosen to reproduce thermodynamic observables rather than microscopic electrostatics. It ensures a balanced interplay between electrostatic and LJ interactions, which is essential for accurately capturing phenomena such as ion partitioning, membrane stability, and protein-lipid interactions.

Bonded Interactions

Bonded interactions are often derived from atomistic simulations or experimental structures. For example, in proteins the secondary structure is frequently maintained fixed with additional constraints or elastic networks, reflecting the limited resolution of the CG representation. As in the atomistic scenario, bonded interactions include bonds, angles, and sometimes dihedrals, described with a functional form similar to that of Atomistic MD. These parameters are often derived from atomistic simulations or experimental structures.

Applications, Limitations and Extensions

The MARTINI force field has been successfully applied to a wide range of systems:

- Lipid bilayers and membrane proteins
- Protein folding and aggregation
- Nanoparticle-membrane interactions
- Polymer and soft-matter systems

Despite its success, MARTINI has known limitations, including:

- Tendency to overestimate aggregation
- Limited resolution for specific interactions
- Challenges in representing secondary structure without additional constraints

Recent developments, such as MARTINI 3, aim to address these issues by improving bead types and overall bead balance [113]. In fact, the MARTINI 3 force field introduces a refined bead classification of the bead interactions, addressing issues such as over-aggregation and limited specificity. It also incorporates more systematic parameterization strategies and broader validation against experimental data [113].

2.4 Introduction to Density Functional Theory

This section is inspired by Ref. [114] and describes the fundamental principles of Density Functional Theory (DFT) calculations, focusing on basis assumptions and approximations. Density Functional Theory is one of the most widely used frameworks for the quantum mechanical description of many-electron systems, ranging from atoms and molecules to solids. Its main goal is to determine the ground-state properties of a system based on the electronic density, rather than the many-body wavefunction, which dramatically reduces computational complexity.

2.4.1 From the Exact Many-Body Problem to the Kohn–Sham Formalism

The exact quantum mechanical description of a system of N electrons and M nuclei is provided by the non-relativistic, time-independent Schrödinger equation:

$$\hat{H}\Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}) = E\Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_I\}), \quad (2.51)$$

where the Hamiltonian contains all kinetic and potential energy terms:

$$\hat{H} = -\sum_{i=1}^N \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_{I=1}^M \frac{\hbar^2}{2M_I} \nabla_I^2 - \sum_{i=1}^N \sum_{I=1}^M \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_I|} + \sum_{i<j}^N \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I<J}^M \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|}. \quad (2.52)$$

Here, the first two terms are the kinetic energies of electrons and nuclei, respectively. The next three terms are the electron–nuclear attraction, electron–electron repulsion, and nuclear–nuclear repulsion. The wavefunction $\Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_I\})$ depends on all $3N + 3M$ spatial coordinates, and in principle also on spin coordinates of electrons. The exact solution of this Schrödinger equation yields all properties of the system; however, the exponential scaling with N and M makes direct solution computationally infeasible for all but the smallest systems. Even representing the wavefunction in a computer for a modest molecule is impossible due to the enormous dimensionality of the Hilbert space.

Given this intractability, approximations are necessary to compute molecular properties. One rigorous alternative is Density Functional Theory. The key insight of DFT is that the ground-state properties of a quantum system are fully determined by the distribution of its electrons in space, rather than by the complete many-body wavefunction. While the full wavefunction $\Psi(\{\mathbf{r}_i\})$ contains all information about correlations and interactions, it depends on $3N$ spatial coordinates and is therefore infeasible to store or manipulate for systems with more than a few electrons. By instead focusing on the electron density:

$$\rho(\mathbf{r}) = N \int |\Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N)|^2 d\mathbf{r}_2 \dots d\mathbf{r}_N, \quad (2.53)$$

which is a function of only three spatial coordinates regardless of the number of electrons, in principle it is possible to capture all ground-state observables. This approach dramatically reduces the complexity of the problem: the high-dimensional wavefunction is replaced by a three-dimensional field, allowing the formal development of rigorous variational principles and functionals that link the density to the total energy of the system. In this way, the electron density becomes the central variable, serving as a practical and computationally tractable surrogate for the full many-body description.

The Hohenberg–Kohn theorems form the foundation of DFT. The **first theorem** states that the ground-state density $\rho_0(\mathbf{r})$ uniquely determines an external potential $V_{\text{ext}}(\mathbf{r})$ up to an additive constant (such external potential) and vice versa. This external potential $V_{\text{ext}}(\mathbf{r})$ originates from the classical interactions between the electrons and the nuclei of the system:

$$V_{\text{ext}}(\mathbf{r}) = - \sum_{I=1}^M \frac{Z_I e^2}{|\mathbf{r} - \mathbf{R}_I|}. \quad (2.54)$$

and its mathematical form in 2.54 corresponds to the third term in equation 2.52. It is a fixed, physically determined potential. In principle, it is the only component of the Hamiltonian that arises directly from the positions and charges of the nuclei. The first Hohenberg–Kohn theorem states that there is a one-to-one mapping between the ground-state density and the external potential (up to a constant). In other words, knowing the electronic density is mathematically sufficient to reconstruct the external potential. This is what allows DFT to be a density-based theory: by knowing the exact density, all the system properties are accessible since all ground-state observables are functionals of $\rho(\mathbf{r})$. The **second theorem** establishes a variational principle: there exists a universal functional $F[\rho]$ such that the total energy

$$E[\rho] = F[\rho] + \int \rho(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) d\mathbf{r} \quad (2.55)$$

is minimized by the exact ground-state density:

$$E_0 = E[\rho_0] = \min_{\rho} E[\rho]. \quad (2.56)$$

The universal functional $F[\rho] = T[\rho] + U[\rho]$ contains the exact kinetic energy $T[\rho]$ of interacting electrons and the electron–electron interaction $U[\rho]$. While formally exact, the explicit dependence of $T[\rho]$ on the density is unknown, which makes direct use of these theorems impractical. In fact, this exact universal functional $F[\rho]$ introduced by Hohenberg and Kohn is unknown and cannot be evaluated directly. Kohn and Sham proposed a solution by introducing a hypothetical, *auxiliary system* of non-interacting electrons that reproduces the exact ground-state density

of the real interacting system. This represents a change of paradigm, from a many-body to a single particle system. In this auxiliary system, the electrons move independently from each other in an effective potential $V_{\text{eff}}(\mathbf{r})$, which incorporates the external potential, the classical electron–electron repulsion (Hartree potential), and a corrective exchange–correlation potential $V_{\text{XC}}(\mathbf{r})$ to account for all many-body effects:

$$V_{\text{eff}}(\mathbf{r}) = \underbrace{V_{\text{ext}}(\mathbf{r})}_{\text{electron-nuclear}} + \underbrace{\int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'}_{\text{Hartree}} + \underbrace{V_{\text{XC}}(\mathbf{r})}_{\text{exchange-correlation}}, \quad (2.57)$$

where

$$V_{\text{XC}}(\mathbf{r}) = \frac{\delta E_{\text{XC}}[\rho]}{\delta \rho(\mathbf{r})}. \quad (2.58)$$

This approach effectively includes the difficult many-body effects into a single term, the exchange–correlation functional $E_{\text{XC}}[\rho]$, while all remaining contributions, including a large portion of the kinetic energy, are treated exactly for the non-interacting system. In this formalism, the total energy functional can be rewritten as:

$$E[\rho] = T_s[\rho] + \int \rho(\mathbf{r}) V_{\text{ext}}(\mathbf{r}) d\mathbf{r} + U[\rho] + E_{\text{XC}}[\rho], \quad (2.59)$$

where $T_s[\rho]$ is the kinetic energy of the non-interacting system and $U[\rho]$ is the classical Hartree energy. The remainder, $E_{\text{XC}}[\rho]$, accounts for the difference between the true kinetic energy and $T_s[\rho]$, as well as all exchange and correlation effects arising from electron–electron interactions. Importantly, $T_s[\rho]$ can be expressed exactly in terms of the Kohn–Sham orbitals $\{\psi_i(\mathbf{r})\}$:

$$T_s[\rho] = -\frac{\hbar^2}{2m} \sum_i^{\text{occ}} \int \psi_i^*(\mathbf{r}) \nabla^2 \psi_i(\mathbf{r}) d\mathbf{r} \quad (2.60)$$

that are solution of the Kohn–Sham equations:

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_{\text{eff}}(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \quad (2.61)$$

where the effective potential $V_{\text{eff}}(\mathbf{r})$ includes the external potential, the Hartree potential, and the exchange–correlation potential, as already stated in equation 2.57. Lastly, the electronic density can be constructed from the occupied orbitals obtained from 2.61:

$$\rho(\mathbf{r}) = \sum_i^{\text{occ}} |\psi_i(\mathbf{r})|^2. \quad (2.62)$$

Intuitively, equations 2.61 and 2.62 can be solved iteratively: starting from an initial guess for $\rho(\mathbf{r})$, one constructs V_{eff} , solves the Kohn–Sham equations to obtain new orbitals, updates the density, and repeats until self-consistency is achieved.

Conceptually, the Kohn–Sham mapping does not correspond to a physical system of non-interacting electrons; rather, it is a mathematical device that reproduces the exact ground-state density. All properties that are functionals of the density, including the total energy, can thus be determined from this auxiliary system. The auxiliary system preserves the complexity of electron–electron interactions entirely within the exchange–correlation functional, which is the only component that requires approximation in practical DFT calculations. This allows one to capture the essential quantum mechanical behavior of electrons while reducing the computational problem to solving a set of self-consistent single-particle equations. The orbitals $\psi_i(\mathbf{r})$ themselves have no direct physical interpretation in terms of the interacting system; their primary role is to generate the correct electron density. The success of this approach lies in the fact that the majority of the kinetic energy is treated exactly through the non-interacting orbitals, which dramatically improves the accuracy compared to approximating the kinetic energy as a functional of the density alone.

In summary, the Kohn–Sham auxiliary system provides a rigorous yet computationally feasible framework to solve the many-electron problem by reducing it to an effective single-particle problem, thus a mean-field approach. It shifts the complexity of interactions into a manageable functional, enabling accurate and widely applicable calculations for atoms, molecules, and solids within the density functional theory framework. Finally, the true goal of DFT is at this point to determine the the exchange–correlation functional.

2.4.2 The Exchange–Correlation Functional and the Local Density Approximation (LDA)

The Local Density Approximation (LDA) is one of the earliest and simplest approximations to the exchange–correlation functional in Kohn–Sham DFT. Its defining feature is that the exchange–correlation energy density at a given point in space is assumed to depend *only on the local value of the electron density* at that point:

$$E_{\text{XC}}^{\text{LDA}}[\rho] = \int \rho(\mathbf{r}) \varepsilon_{\text{XC}}^{\text{hom}}(\rho(\mathbf{r})) d\mathbf{r}. \quad (2.63)$$

Here, $\varepsilon_{\text{XC}}^{\text{hom}}(\rho)$ is the exchange–correlation energy per electron of a uniform electron gas with density ρ . The adjective “*local*” arises from this pointwise treatment: the functional at $\mathbf{r}$ depends only on $\rho(\mathbf{r})$, and *not* on gradients, Laplacians, or values of the density at neighboring points. In other words, LDA assumes that the electron density is slowly varying enough locally that it

can be treated as homogeneous in an infinitesimal volume element.

The corresponding exchange–correlation potential, which enters the Kohn–Sham equations, is obtained as the functional derivative of the energy with respect to the density:

$$V_{\text{XC}}^{\text{LDA}}(\mathbf{r}) = \frac{\delta E_{\text{XC}}^{\text{LDA}}[\rho]}{\delta \rho(\mathbf{r})} = \varepsilon_{\text{XC}}^{\text{hom}}(\rho(\mathbf{r})) + \rho(\mathbf{r}) \left. \frac{d\varepsilon_{\text{XC}}^{\text{hom}}}{d\rho} \right|_{\rho=\rho(\mathbf{r})}. \quad (2.64)$$

The locality of $V_{\text{XC}}^{\text{LDA}}(\mathbf{r})$ is a direct consequence: the potential at a point depends only on the density at that point.

When implementing DFT and in a computational code, the choice of basis set affects how orbitals and densities are represented, but does not alter the formal locality of LDA. Two common types of basis sets are used.

Local (atomic orbital) basis sets In a local basis, the Kohn–Sham orbitals $\psi_i(\mathbf{r})$ are expanded as linear combinations of localized functions centered on atoms:

$$\psi_i(\mathbf{r}) = \sum_{\mu} c_{\mu i} \phi_{\mu}(\mathbf{r} - \mathbf{R}_{\mu}), \quad (2.65)$$

where ϕ_{μ} are typically Gaussian-type orbitals or numerical atomic orbitals. In this case:

- The electron density $\rho(\mathbf{r}) = \sum_i^{\text{occ}} |\psi_i(\mathbf{r})|^2$ is naturally localized around atoms.
- The evaluation of the local LDA potential $V_{\text{XC}}^{\text{LDA}}(\mathbf{r})$ is straightforward pointwise on a grid or at the atomic centers.
- The locality of LDA works efficiently with local basis sets, since the potential and density share the same spatial locality.

Plane-wave basis sets In a plane-wave basis, orbitals are expanded as delocalized functions over the simulation cell:

$$\psi_i(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_{i\mathbf{G}} e^{i\mathbf{G} \cdot \mathbf{r}}, \quad (2.66)$$

where $\mathbf{G}$ are reciprocal lattice vectors and Ω is the cell volume. In this representation:

- The density $\rho(\mathbf{r})$ is obtained by summing the squared magnitudes of orbitals, which can be efficiently transformed between real and reciprocal space using Fast Fourier Transforms (FFT).
- The local nature of LDA is realized in *real space*, even though the orbitals are delocalized. The LDA potential is applied pointwise on a real-space grid, then transformed back to reciprocal space to act on plane-wave coefficients.

- Despite the delocalization of the basis, LDA remains a local approximation because the potential at a given grid point depends only on the density at that point.

In essence, LDA is “local” because it treats exchange–correlation effects at a point in space independently of the surrounding density. This locality is naturally aligned with local basis sets but can also be implemented efficiently in plane-wave codes using grid-based representations. The choice of basis affects computational efficiency and convergence, but does not alter the formal definition of the approximation.

2.4.3 Algorithm of the DFT Procedure

In summary, a DFT calculation proceeds as follows:

1. Make an initial guess for the electron density $\rho(\mathbf{r})$.
2. Construct the effective Kohn–Sham potential $V_{\text{eff}}(\mathbf{r})$ using the guessed density.
3. Solve the Kohn–Sham equations to obtain orbitals $\psi_i(\mathbf{r})$ and their energies ϵ_i .
4. Compute a new density from the orbitals: $\rho(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2$.
5. Equations in point 3 and 4 are then solved self-consistently to obtain the ground-state density and energy. A check for convergence is then performed, and if not converged, update the density and repeat steps 2–4.
6. Once converged, compute the ground-state total energy using the same framework shown in equation 2.59. Practically, this equation is solved *via* the following Kohn–Sham DFT framework. In equation 2.59, $T_s[\rho]$ is the kinetic energy of the non-interacting Kohn–Sham system, $V_{\text{ext}}(\mathbf{r})$ is the external potential, $U[\rho]$ is the Hartree energy, and $E_{\text{XC}}[\rho]$ is the exchange–correlation functional. Specifically, the Hartree energy is:

$$U[\rho] = \frac{1}{2} \int \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \quad (2.67)$$

The Kohn–Sham orbitals satisfy equation 2.61, so summing over the occupied eigenvalues ϵ_i gives $\sum_i \epsilon_i = T_s + \int \rho V_{\text{ext}} d\mathbf{r} + \int \rho V_H d\mathbf{r} + \int \rho V_{\text{XC}} d\mathbf{r}$, while the Hartree potential satisfies $\int \rho(\mathbf{r}) V_H(\mathbf{r}) d\mathbf{r} = 2U[\rho]$. Removing the double counting terms leads to the following definition to the total energy functional:

$$E[\rho] = \sum_i \epsilon_i - U[\rho] + E_{\text{XC}}[\rho] - \int \rho(\mathbf{r}) V_{\text{XC}}(\mathbf{r}) d\mathbf{r}. \quad (2.68)$$

This framework provides a computationally efficient approach to accurately model the electronic structure of complex systems while retaining the essential quantum mechanical interactions.

2.5 Introduction to Real Time, Time Dependent Density Functional Theory

The purpose of this chapter is to extend the framework of Density Functional Theory (DFT) to the time-dependent regime, enabling the description of electronic dynamics under external perturbations. While the previous chapter introduced the theoretical structure of ground-state DFT—including the Hohenberg–Kohn theorems, the Kohn–Sham construction, and practical approximations such as the Local Density Approximation (LDA)—those concepts describe only stationary electronic ground states. However, many physical phenomena require a time-dependent treatment, including: optical absorption and emission, electron dynamics in strong laser fields, ultrafast charge transfer, nonadiabatic molecular processes. Photoexcited molecular dynamics Time-Dependent Density Functional Theory (TDDFT) provides a formally exact framework for such phenomena by extending the density-functional paradigm to the time-dependent many-electron Schrödinger equation. Thus, the purpose of the current section is to extend the Kohn–Sham framework to the description of electronic systems evolving in time.

Among TDDFT approaches, two main computational strategies exist: Linear-response TDDFT (frequency-domain methods) and Real-Time TDDFT (RT-TDDFT). This chapter focuses primarily on RT-TDDFT, where the time-dependent Kohn–Sham equations are propagated explicitly in time, allowing simulation of electronic dynamics directly. Moreover, this approach allows to simulate the full dynamical evolution of an interacting electronic system under external perturbations, like lasers or external fields (this will become relevant in the next chapters). Finally, the coupling of RT-TDDFT with nuclear motion will be briefly introduced, leading to Ehrenfest dynamics, which provides a first-principles framework for simulating the time evolution of both electrons and nuclei.

2.5.1 Foundations of Time-Dependent Density Functional Theory: the Time-Dependent Kohn–Sham Equations

Time-Dependent Density Functional Theory generalizes the Hohenberg–Kohn framework to systems evolving in time. Consider an interacting many-electron system subject to a time-dependent external potential $V_{\text{ext}}(\mathbf{r}, t)$. The full many-body wavefunction satisfies the time-dependent Schrödinger equation:

$$i \frac{\partial}{\partial t} \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N, t) = \hat{H}(t) \Psi(\mathbf{r}_1, \dots, \mathbf{r}_N, t), \quad (2.69)$$

where Hamiltonian now explicitly includes a time-dependent external potential:

$$\hat{H}(t) = \sum_i \left(-\frac{1}{2} \nabla_i^2 + V_{\text{ext}}(\mathbf{r}_i, t) \right) + \sum_{i < j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (2.70)$$

The fundamental theorem underlying TDDFT was established by Runge and Gross. The Runge–Gross theorem states that, given an initial many-body state Ψ_0 , the time-dependent density $n(\mathbf{r}, t)$ uniquely determines the time-dependent external potential $v_{\text{ext}}(\mathbf{r}, t)$ up to a purely time-dependent function. Consequently, all observables of the system become functionals of the density and the initial state. As in ground-state DFT, the interacting system can be mapped onto a fictitious system of non-interacting electrons that reproduces the exact time-dependent density. The auxiliary system satisfies the time-dependent Kohn–Sham equations:

$$i \frac{\partial}{\partial t} \psi_i(\mathbf{r}, t) = \left[-\frac{1}{2} \nabla^2 + V_{\text{KS}}(\mathbf{r}, t) \right] \psi_i(\mathbf{r}, t), \quad (2.71)$$

and the time-dependent density is given by:

$$\rho(\mathbf{r}, t) = \sum_i^{\text{occ}} |\psi_i(\mathbf{r}, t)|^2, \quad (2.72)$$

where the time-dependent Kohn–Sham potential has the structure

$$V_{\text{KS}}(\mathbf{r}, t) = V_{\text{ext}}(\mathbf{r}, t) + V_{\text{H}}(\mathbf{r}, t) + V_{\text{XC}}(\mathbf{r}, t). \quad (2.73)$$

The Hartree potential retains the same form as in ground-state theory:

$$V_{\text{H}}(\mathbf{r}, t) = \int \frac{\rho(\mathbf{r}', t)}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}'. \quad (2.74)$$

The exchange-correlation potential is now time dependent:

$$V_{\text{XC}}(\mathbf{r}, t) = \frac{\delta E_{\text{XC}}[\rho]}{\delta \rho(\mathbf{r}, t)}, \quad (2.75)$$

where E_{XC} is the exchange-correlation energy functional. In a practical implementations one commonly adopts the adiabatic approximation:

$$V_{\text{XC}}(\mathbf{r}, t) \approx V_{\text{XC}}^{\text{GS}}[\rho(\mathbf{r}, t)], \quad (2.76)$$

which evaluates the exchange-correlation potential using the instantaneous density as if the system were in its ground state.

2.5.2 Real-Time Propagation in Time Dependent Density Functional Theory

In the real-time (RT) formulation of TDDFT, the time-dependent Kohn–Sham equations are solved directly by propagating the orbitals forward in time. Formally, the evolution operator $\hat{H}(t)$ satisfies:

$$\psi_i(t + \Delta t) = \hat{H}(t + \Delta t, t)\psi_i(t) \quad (2.77)$$

with:

$$\hat{H}(t + \Delta t, t) = \exp \left[-i \int_t^{t+\Delta t} \hat{H}_{KS}(t') dt' \right], \quad (2.78)$$

where $\hat{H}_{KS}(t')$ is the Hamiltonian of the time-dependent Kohn–Sham equation 2.71. However, for sufficiently small time steps, the Kohn–Sham time-dependent Hamiltonian can be approximated as constant, rewriting equation 2.77 as follows:

$$\psi_i(t + \Delta t) \approx \exp[-i\hat{H}_{KS}(t)\Delta t]\psi_i(t). \quad (2.79)$$

Various numerical propagation schemes can be used in RT-TDDFT:

- Crank–Nicolson propagator
- split-operator methods
- enforced time-reversal symmetry schemes
- Runge–Kutta integration

A commonly used unitary scheme is the Crank–Nicolson propagator. The Crank–Nicolson propagator can be derived by applying the trapezoidal rule to the time derivative of the wavefunction. Starting from:

$$\frac{\partial \psi_i}{\partial t} = -i\hat{H}_{KS}\psi_i, \quad (2.80)$$

this derivative can be approximated at the midpoint between t and $t + \Delta t$:

$$\frac{\psi_i(t + \Delta t) - \psi_i(t)}{\Delta t} = -\frac{i}{2}\hat{H}_{KS}[\psi_i(t + \Delta t) + \psi_i(t)]. \quad (2.81)$$

Rearranging this equation gives:

$$\psi_i(t + \Delta t) - \psi_i(t) = -\frac{i\Delta t}{2}\hat{H}_{KS}[\psi_i(t + \Delta t) + \psi_i(t)]. \quad (2.82)$$

Collecting the terms containing $\psi_i(t + \Delta t)$ yields:

$$\left(1 + \frac{i\Delta t}{2}\hat{H}_{KS}\right)\psi_i(t + \Delta t) = \left(1 - \frac{i\Delta t}{2}\hat{H}_{KS}\right)\psi_i(t). \quad (2.83)$$

This is the Crank–Nicolson propagation equation. Formally solving for $\psi_i(t + \Delta t)$ gives back equation

$$\psi_i(t + \Delta t) = \left(1 + \frac{i\Delta t}{2}\hat{H}_{KS}\right)^{-1} \left(1 - \frac{i\Delta t}{2}\hat{H}_{KS}\right) \psi_i(t). \quad (2.84)$$

This operator can be interpreted as a rational approximation to the exponential propagator:

$$\exp(-i\hat{H}_{KS}\Delta t) \approx \frac{1 - \frac{i\Delta t}{2}\hat{H}_{KS}}{1 + \frac{i\Delta t}{2}\hat{H}_{KS}}, \quad (2.85)$$

and such approximation corresponds to the second-order Padé approximation of the exponential function. The Crank–Nicolson scheme possesses several important numerical properties that make it particularly suitable for quantum dynamics. First, it is **second-order accurate in time**. The local truncation error scales as $\mathcal{O}(\Delta t^3)$, while the global error scales as $\mathcal{O}(\Delta t^2)$. Second, the propagator is **unitary**, meaning the norm of each orbital is preserved during propagation. Third, the method is **numerically stable**. Unlike explicit time integration schemes, the Crank–Nicolson algorithm remains stable even for relatively large time steps. For these reasons it has historically been one of the most widely used propagation schemes in early TDDFT implementations, although modern large-scale simulations often employ more advanced exponential propagators or enforced time-reversal symmetry algorithms. Nevertheless, the Crank–Nicolson method remains conceptually important because it clearly illustrates how the time-dependent Kohn–Sham equations can be integrated while preserving the fundamental properties of quantum evolution.

Lastly, it is important to note that the Kohn–Sham Hamiltonian depends on the density, which itself depends on the propagated orbitals. Consequently, the propagation of equations 2.72 and 2.84 must remain self-consistent. A typical time step proceeds as follows:

1. Compute the density $\rho(\mathbf{r}, t)$ from the current orbitals.
2. Construct the Kohn–Sham Hamiltonian $\hat{H}_{KS}[\rho(r, t)]$.
3. Apply the Crank–Nicolson propagation to obtain tentative orbitals at $t + \Delta t$.
4. Recompute the density $\rho(\mathbf{r}, t + \Delta t)$.
5. Update the Hamiltonian if necessary.

More sophisticated algorithms evaluate the Hamiltonian at the midpoint time $t + \Delta t/2$ to improve accuracy.

2.5.3 Coupled Electron–Nuclear Dynamics: TDDFT and the Ehrenfest Approximation

Thus far, nuclei have always been considered as fixed. This approximation, often referred to as the *Born–Oppenheimer approximation* (see section 2.1.2), is extremely useful for computing electronic ground states and excited states at fixed nuclear geometries. However, many physical and chemical processes involve a strong coupling between electronic motion and nuclear motion. Examples include photochemical reactions, laser-induced molecular dynamics, charge transfer processes, vibronic coupling, and nonadiabatic molecular rearrangements. In order to properly model the dynamics of a molecule from first principles, it is therefore necessary to treat both electrons and nuclei as dynamical degrees of freedom. In principle, the correct description would require solving the full time-dependent Schrödinger equation for both electrons and nuclei simultaneously. Given the electronic coordinates collectively as $\mathbf{r}$ and the nuclear coordinates as $\mathbf{R}$, the full wavefunction of the system is $\Psi(\mathbf{r}, \mathbf{R}, t)$, which evolves according to:

$$i \frac{\partial}{\partial t} \Psi(\mathbf{r}, \mathbf{R}, t) = \hat{H}(\mathbf{r}, \mathbf{R}) \Psi(\mathbf{r}, \mathbf{R}, t). \quad (2.86)$$

The total Hamiltonian of the coupled system is:

$$\hat{H} = \hat{T}_e + \hat{T}_N + \hat{V}_{ee} + \hat{V}_{eN} + \hat{V}_{NN}, \quad (2.87)$$

where $\hat{T}_e$ and $\hat{T}_N$ are the kinetic energy operators for electrons and nuclei respectively, while $\hat{V}_{ee}$, $\hat{V}_{eN}$, and $\hat{V}_{NN}$ represent the electron–electron, electron–nucleus, and nucleus–nucleus Coulomb interactions. Solving this equation exactly for realistic molecular systems is computationally impossible due to the enormous dimensionality of the nuclear configuration space. For this reason, practical first-principles molecular dynamics approaches rely on mixed quantum–classical approximations in which the electrons are treated quantum mechanically while the nuclei are described by classical mechanics. One of the simplest and most widely used approaches of this type is the *Ehrenfest dynamics* method.

The Ehrenfest approximation can be derived from Ehrenfest’s theorem, which relates the time evolution of the expectation value of an observable to the corresponding classical equation of motion. Applying this principle to the nuclear coordinates leads to classical equations of motion for the nuclei in which the forces are obtained as quantum expectation values over the electronic wavefunction. Within this framework, the nuclei are treated as classical particles with positions $\mathbf{R}_I(t)$ and masses M_I , whose dynamics is governed by Newton’s equations:

$$M_I \frac{d^2 \mathbf{R}_I}{dt^2} = \mathbf{F}_I. \quad (2.88)$$

The force acting on nucleus I arises from the instantaneous electronic structure of the system and is given by the expectation value of the gradient of the Hamiltonian with respect to the nuclear coordinates:

$$\mathbf{F}_I = - \left\langle \Psi(t) \left| \frac{\partial \hat{H}}{\partial \mathbf{R}_I} \right| \Psi(t) \right\rangle. \quad (2.89)$$

This expression is essentially a manifestation of the Hellmann–Feynman theorem generalized to time-dependent quantum states. It states that the force on each nucleus can be obtained by differentiating the electronic Hamiltonian with respect to the nuclear position and evaluating the expectation value over the instantaneous electronic state. Physically, this means that the nuclei move under the influence of the average electronic force generated by the evolving electronic density.

In the context of Time-Dependent Density Functional Theory, the full many-electron wavefunction $\Psi(t)$ is replaced by the set of time-dependent Kohn–Sham orbitals $\psi_i(\mathbf{r}, t)$. Since the electronic density is reconstructed from these orbitals as in equation 2.72, the forces on the nuclei can be expressed directly in terms of the Kohn–Sham Hamiltonian. Within TDDFT formalism, the force on nucleus I can therefore be written as:

$$\mathbf{F}_I = - \sum_i \left\langle \psi_i \left| \frac{\partial \hat{H}_{KS}}{\partial \mathbf{R}_I} \right| \psi_i \right\rangle. \quad (2.90)$$

The derivative of the Kohn–Sham Hamiltonian with respect to the nuclear coordinates arises primarily from the dependence of the external potential on the nuclear positions:

$$V_{\text{ext}}(\mathbf{r}, t) = - \sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I(t)|}, \quad (2.91)$$

where Z_I is the charge of nucleus I . Consequently, the nuclear forces originate from the interaction between the electronic density and the nuclear Coulomb potential, as well as from the direct nucleus–nucleus repulsion. In practice, the total force on a nucleus can be decomposed into two contributions:

$$\mathbf{F}_I = \mathbf{F}_I^{eN} + \mathbf{F}_I^{NN}. \quad (2.92)$$

The electron–nucleus force arises from the interaction with the electronic density:

$$\mathbf{F}_I^{eN} = - \int \rho(\mathbf{r}, t) \frac{\partial V_{\text{ext}}(\mathbf{r}, \mathbf{R})}{\partial \mathbf{R}_I} d\mathbf{r}, \quad (2.93)$$

while the nucleus–nucleus force is given by the classical Coulomb interaction between nuclei:

$$\mathbf{F}_I^{NN} = - \sum_{J \neq I} \frac{\partial}{\partial \mathbf{R}_I} \left(\frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} \right). \quad (2.94)$$

The key feature of Ehrenfest dynamics is that the electronic and nuclear degrees of freedom evolve simultaneously and influence each other in a self-consistent way, giving rise to a self-consistent electron–nuclear dynamics. At each time step of the simulation: the electronic wavefunctions are propagated according to the time-dependent Kohn–Sham equations using the current nuclear positions. Then, the electronic density is updated from the propagated orbitals, and the forces on the nuclei are computed from that electronic density. Lastly, the nuclear positions and velocities are updated using classical molecular dynamics integration schemes such as the velocity-Verlet algorithm. Thus, the nuclear motion modifies the external potential experienced by the electrons, which in turn affects the subsequent electronic evolution. This feedback loop naturally captures the coupling between electronic excitations and nuclear motion.

Ehrenfest dynamics can be interpreted as a mean-field approach to nonadiabatic dynamics. The electrons evolve according to the time-dependent quantum Hamiltonian generated by the moving nuclei, while the nuclei feel the average force produced by the electronic density. However, the Ehrenfest approximation also has important limitations. In particular, it treats nuclear motion as classical and does not account for nuclear quantum effects such as tunneling or zero-point motion. Moreover, because the nuclei feel an averaged electronic force, the method cannot correctly describe situations where the system evolves along multiple potential energy surfaces simultaneously, such as in strong nonadiabatic transitions. Despite these limitations, Ehrenfest dynamics combined with RT-TDDFT still remains one of the most widely used approaches for simulating coupled electron–nuclear dynamics in molecules and condensed matter systems, especially in the context of ultrafast phenomena and laser-driven processes.

2.5.4 Algorithm for First-Principles Molecular Dynamics with RT-TDDFT and Ehrenfest

To sum up, a typical, general RT-TDDFT Ehrenfest dynamics simulation scheme proceeds as follows:

1. Initialize nuclear positions $\mathbf{R}_I(0)$ and velocities.
2. Solve the ground-state Kohn–Sham equations to obtain the initial orbitals $\psi_i(\mathbf{r}, 0)$.
3. Construct the initial density: $\rho(\mathbf{r}, 0) = \sum_i |\psi_i(\mathbf{r}, 0)|^2$.
4. Build the time-dependent Kohn–Sham Hamiltonian.

5. Propagate the Kohn–Sham orbitals for a time step Δt using a unitary propagator.
6. Compute the updated density $\rho(\mathbf{r}, t)$.
7. Evaluate the forces on the nuclei.
8. Integrate the classical equations of motion: $M_I \frac{d^2 \mathbf{R}_I}{dt^2} = \mathbf{F}_I$.
9. Update nuclear positions and velocities.
10. Recompute the external potential from the new nuclear coordinates.
11. Repeat the propagation cycle.

This combined scheme allows the simulation of electronic excitations, energy transfer, charge migration, and nonadiabatic molecular motion directly from first principles. RT-TDDFT coupled with Ehrenfest dynamics therefore provides a powerful framework for studying ultrafast processes in molecules and condensed matter systems, bridging electronic structure theory and dynamical simulations of complex quantum systems.

2.6 Density Functional Theory and Linear Optical Response

This section focuses on a specific application of Density Functional Theory on the modeling of a very well-known phenomena, namely optical absorption of bulk materials (like silica matrices) in the framework of the linear optical response approximation. This section is essential to understand results from Chapter 5, and is inspired by Ref. [114].

2.6.1 Introduction to Optical Absorption and Linear Response Theory

Optical absorption is the process by which a material absorbs energy from an incident electromagnetic field, leading to electronic excitations within the probed system. When a certain light of frequency ω interacts with a solid, photons carrying an energy $\hbar\omega$ can be absorbed if their energy matches the difference between two allowed electronic states. In semiconductors and insulators, this typically corresponds to the excitation of an electron from a valence band state to an empty conduction band state. As a consequence of this process, the intensity of the electromagnetic wave of the incoming light field decreases as it propagates through the material.

Thus, the optical absorption properties of bulk insulating materials originate from the interaction between electromagnetic radiation and their electronic structures. Moreover, such interaction is fundamentally governed by the response of the electronic charge density to an external time-dependent perturbation. As stated earlier, optical absorption arises when photons

promote electrons from occupied states to higher-energy unoccupied states, thereby transferring energy from the electromagnetic field to the electronic system. In the regime relevant to most spectroscopic experiments, the external electromagnetic field is sufficiently weak that the response of the material can be treated within the framework of linear response theory. The term *linear response* refers to the approximation in which the induced polarization of the material depends linearly on the applied external electric field. In the context of optical absorption, the incident electromagnetic field is typically weak compared to the internal electronic interactions of the system, so higher-order effects such as nonlinear polarization, multi-photon absorption, or field-induced changes in the electronic structure are negligible. Under these conditions, the induced polarization $\mathbf{P}(\omega)$ can be expressed as a first-order expansion in the applied field $\mathbf{E}(\omega)$:

$$\mathbf{P}(\omega) = \chi(\omega)\mathbf{E}(\omega), \quad (2.95)$$

where $\chi(\omega)$ is the frequency-dependent electric susceptibility tensor, that characterizes how the electronic charge distribution of the material reorganizes in response to the perturbing field. This linear relation implies that the material's response at frequency ω is directly proportional to the perturbing field at the same frequency, allowing the use of linear response theory.

In fact, linear response provides a rigorous framework to compute macroscopic optical properties, such as the dielectric function and absorption coefficient, from microscopic electronic transitions while ensuring that causality and energy conservation are satisfied. From the $\chi(\omega)$, it is possible to compute the macroscopic dielectric function as follows:

$$\varepsilon(\omega) = 1 + 4\pi\chi(\omega), \quad (2.96)$$

that governs the optical properties of condensed matter systems. In fact, at the macroscopic level, the interaction between electromagnetic radiation and matter is described by such frequency-dependent complex dielectric function that can be decomposed as:

$$\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega). \quad (2.97)$$

The real part $\varepsilon_1(\omega)$ describes the dispersive response of the medium, which determines how the phase velocity of light is modified inside the material, while the imaginary part $\varepsilon_2(\omega)$ accounts for the dissipative processes responsible for the absorption of electromagnetic energy. Therefore, the optical absorption spectrum of a material is directly related to the imaginary part of the dielectric function. In particular, the absorption coefficient $\alpha(\omega)$, which is commonly measured

in optical spectroscopy experiments, is related to the dielectric function through

$$\alpha(\omega) = \frac{\sqrt{2}\omega}{c} \left[\sqrt{\varepsilon_1^2(\omega) + \varepsilon_2^2(\omega)} - \varepsilon_1(\omega) \right]^{1/2}. \quad (2.98)$$

This relation connects the macroscopic attenuation of light in a material ($\alpha(\omega)$) to the complex dielectric function $\varepsilon(\omega)$, which is determined by the electronic transitions between occupied and unoccupied states of the system. Consequently, the absorption coefficient $\alpha(\omega)$ encodes information about the electronic band structure and the allowed optical transitions, establishing a connection between experimentally measured optical spectra and the microscopic electronic structure of the material.

From a microscopic point of view, the optical response originates from electronic transitions between quantum states of the system. In a periodic crystal, the electronic states are described by Bloch functions $\psi_{n\mathbf{k}}(\mathbf{r})$, which are labeled by a band index n and a wavevector $\mathbf{k}$ within the first Brillouin zone. The interaction with an external electromagnetic field can induce transitions from occupied valence states to unoccupied conduction states, leading to the absorption of photons. Within the electric dipole approximation (which assumes that the wavelength of light is much larger than the unit cell) and the independent particle picture (see Section 2.6.2), the imaginary part of the dielectric tensor $\varepsilon_2^{\alpha\beta}(\omega)$ can be expressed as a sum over interband transitions between occupied and unoccupied electronic states throughout the Brillouin zone:

$$\varepsilon_2^{\alpha\beta}(\omega) = \frac{8\pi^2 e^2}{\Omega} \sum_{v,c} \int_{\text{BZ}} \frac{d\mathbf{k}}{(2\pi)^3} |\langle \psi_{c\mathbf{k}} | \hat{v}_\alpha | \psi_{v\mathbf{k}} \rangle|^2 \delta(E_{c\mathbf{k}} - E_{v\mathbf{k}} - \hbar\omega), \quad (2.99)$$

where:

- v and c denote valence and conduction band Bloch states, respectively;
- Ω is the volume of the unit cell; note that optical experiments probe a macroscopic response, whereas first-principles calculations yield a microscopic, spatially varying dielectric function. For this reason, a spatial average of the dielectric tensor is performed in order to obtain a macroscopic quantity that can be directly related to experimentally measurable optical properties. This spatially-averaged microscopic imaginary dielectric function $\varepsilon_2^{\alpha\beta}(\omega)$ captures the collective response of the electrons and can be directly related to optical absorption, thus to macroscopic properties, like $\alpha(\omega)$.
- $\hat{v}_\alpha$ is the component of the velocity operator along the Cartesian direction α ;
- the integral is performed over the first Brillouin zone (BZ). In a periodic crystal, the first Brillouin zone (BZ) is the fundamental region of the reciprocal space, containing all

distinct wavevectors $\mathbf{k}$ that describe the electronic states. It can be defined as:

$$\text{BZ} = \left\{ \mathbf{k} \in \mathbb{R}^3 \mid \text{closest to the origin in the reciprocal lattice} \right\} \quad (2.100)$$

Or, using the reciprocal lattice vectors $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ and the direct lattice vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, the Brillouin zone can be written as:

$$\text{BZ} = \left\{ \mathbf{k} \in \mathbb{R}^3 \mid |\mathbf{k} \cdot \mathbf{a}_i| \leq \frac{|\mathbf{b}_i|}{2}, i = 1, 2, 3 \right\} \quad (2.101)$$

In equation 2.99, the integral over the Brillouin zone sums contributions from all unique $\mathbf{k}$ -points. Its physical meaning can be described as follows: electrons in a bulk are described by Bloch states $|\psi_{n\mathbf{k}}\rangle$ labeled by $\mathbf{k}$. Due to periodicity, wavevectors outside the first BZ are equivalent to some $\mathbf{k}$ inside it. Therefore, integrating over the first BZ captures all distinct electronic contributions to optical absorption. Moreover, For each $\mathbf{k}$ in the BZ, the formula sums contributions from all possible valence-to-conduction band transitions. Physically, this accounts for how electrons throughout the bulk respond to light, not just at one point in reciprocal space. The peaks in the imaginary part $\varepsilon_2(\omega)$ arise from transitions where the matrix elements are strong. The Brillouin zone defines the map over which these electronic states exist, so its shape and symmetry affect the optical properties.

- the Dirac delta function $\delta(E_{c\mathbf{k}} - E_{v\mathbf{k}} - \hbar\omega)$ enforces energy conservation, allowing only transitions for which the photon energy $\hbar\omega$ matches the energy difference between initial and final states, so only transitions with energy difference $\hbar\omega$. In practical implementations, the delta function is replaced by a smearing function, typically a Lorentzian or Gaussian of finite width η , in order to mimic lifetime broadening and facilitate numerical integration:

$$\delta(E_{c\mathbf{k}} - E_{v\mathbf{k}} - \hbar\omega) \rightarrow \frac{1}{\pi} \frac{\eta}{(E_{c\mathbf{k}} - E_{v\mathbf{k}} - \hbar\omega)^2 + \eta^2}. \quad (2.102)$$

This broadening parameter plays a crucial role in shaping the computed absorption spectra, determining the width and smoothness of the spectral features.

Equation 2.99 reveals that the optical absorption spectrum is governed by two key factors: the joint density of states (JDOS) $\sum_{v,c} \int_{\text{BZ}} \frac{d\mathbf{k}}{(2\pi)^3} \delta(E_{c\mathbf{k}} - E_{v\mathbf{k}} - \hbar\omega)$, which counts the number of pairs of occupied and unoccupied states separated by an energy $\hbar\omega$; the dipole (or velocity) matrix elements, $\langle \psi_{c\mathbf{k}} | \hat{v}_\alpha | \psi_{v\mathbf{k}} \rangle$, which determine the probability amplitude of the transition. Physically, even if many electronic states are available at a given energy difference, a transition may be weak or forbidden if the corresponding matrix element is small due to symmetry or selection rules. Therefore, the absorption spectrum is shaped by both the electronic structure and the allowed

optical transitions, providing a direct connection between the microscopic quantum states of the material and the experimentally observed optical properties.

2.6.2 Density Functional Theory, Linear Optical Response and Absorption Spectra

Determining the transition matrix elements $\langle \psi_{c\mathbf{k}} | \hat{v}_\alpha | \psi_{v\mathbf{k}} \rangle$ is the goal of any theoretical approach aiming at evaluating the Linear Optical Response of a material. In fact, as stated in equations 2.98 and 2.99, obtaining the imaginary part of the dielectric function leads to the determination of the relative absorption coefficient.

Density Functional Theory can provide a first-principles framework that allows to compute these quantities starting from the atomic structure of the system. As introduced previously, DFT reformulates the interacting many-electron problem in terms of a set of auxiliary non-interacting particles that evolve in an effective potential. The electronic states $\psi(\mathbf{r})$ and the relative eigenvalues ϵ are obtained by solving the Kohn-Sham equations (2.61). Retaining the Bloch notation, the electron density is reconstructed from the Kohn-Sham orbitals according to $\rho(\mathbf{r}) = \sum_{n\mathbf{k}} f_{n\mathbf{k}} |\psi_{n\mathbf{k}}(\mathbf{r})|^2$ (equal to equation 2.62), where $f_{n\mathbf{k}}$ represents the occupation factor of each state. Once the self-consistent solution of equations 2.61 and 2.62 is obtained, the eigenvalues $\epsilon_{n\mathbf{k}}$ and eigenfunctions $\psi_{n\mathbf{k}}$ provide the fundamental ingredients required for calculating the optical response of the material.

In this context, it is essential to introduce the independent particle approximation (IPA). IPA refers to the assumption that the excited electrons and the holes they leave behind behave as non-interacting particles, even during an optical transition. More precisely, within standard DFT the many-body problem of interacting electrons is mapped onto a system of non-interacting Kohn-Sham particles moving in an effective potential. When computing optical properties in the IPA, this idea is extended to the excitation process itself, assuming that an electron can be promoted from a valence state $\psi_{v\mathbf{k}}$ to a conduction state $\psi_{c\mathbf{k}}$ without any additional interaction with other electrons beyond what is already included in the ground-state effective potential. Formally, this implies that: (i) the excited state is approximated as a simple single-particle transition from valence to conduction states; (ii) the electron and the hole do not interact with each other after the excitation; (iii) many-body effects such as electron-hole attraction (excitons) and dynamical screening are neglected. From a physical point of view, IPA corresponds to a picture in which each electron responds to the external electromagnetic field as if it were moving independently, without feeling the presence of the hole it leaves behind or the correlated motion of the other electrons.

Within this approximation, the optical transition matrix elements entering the dielectric function can be expressed using the momentum operator $\hat{\mathbf{p}}$. The transition probability between

a valence state v and a conduction state c is determined by

$$M_{cv}(\mathbf{k}) = \langle \psi_{c\mathbf{k}} | \hat{\mathbf{p}} | \psi_{v\mathbf{k}} \rangle, \quad (2.103)$$

and using these matrix elements, the imaginary part of the dielectric function can also be written as:

$$\varepsilon_2(\omega) = \frac{4\pi^2 e^2}{m^2 \omega^2 \Omega} \sum_{v,c} \int_{\text{BZ}} \frac{d\mathbf{k}}{(2\pi)^3} |M_{cv}(\mathbf{k})|^2 \delta(E_{c\mathbf{k}} - E_{v\mathbf{k}} - \hbar\omega). \quad (2.104)$$

Here, as before the integral runs over the first Brillouin zone, ensuring that all distinct electronic states are included; the Dirac delta function enforces energy conservation, allowing only transitions where the photon energy $\hbar\omega$ matches the difference between conduction and valence band energies. Note that the squared matrix element $|M_{cv}(\mathbf{k})|^2$ accounts for the strength of coupling between electrons and the electromagnetic field.

The real part of the dielectric function, $\varepsilon_1(\omega)$, which describes the dispersive response of the material to light of angular frequency ω , can be obtained from the imaginary part $\varepsilon_2(\omega)$ using the Kramers-Kronig relation:

$$\varepsilon_1(\omega) = 1 + \frac{2}{\pi} \mathcal{P} \int_0^\infty \frac{\omega' \varepsilon_2(\omega')}{\omega'^2 - \omega^2} d\omega', \quad (2.105)$$

where:

- ω is the angular frequency of the incident light.
- $\varepsilon_2(\omega')$ is the imaginary part of the dielectric function at frequency ω' .
- $\mathcal{P}$ denotes the principal value of the integral, which properly handles the singularity at $\omega' = \omega$.

Physically, this relation reflects the fact that the absorption spectrum $\varepsilon_2(\omega)$ determines the dispersive behavior $\varepsilon_1(\omega)$, ensuring a causal response of the material to the electromagnetic field.

From a computational perspective, these two quantities are typically computed using plane-wave implementations of DFT: within this framework, the Kohn-Sham orbitals are expanded in a plane-wave basis set, as sums of reciprocal lattice vectors $\mathbf{G}$, as reported in equation 2.66. A dense sampling of the Brillouin zone is needed, and convergence with respect to the $\mathbf{k}$ -point mesh, the number of empty conduction bands included in the summation, and the plane-wave cutoff are essential to obtain reliable results. In addition, the interaction between valence electrons and ionic cores is typically described using pseudopotentials, allowing to remove the chemically inert core electrons and reduce the computational cost of the calculations. After a self-consistent

field calculation determines the ground-state electronic density, the Kohn–Sham eigenvalues and wavefunctions can be used in a subsequent linear response calculation to evaluate the dielectric tensor $\varepsilon_2^{\alpha\beta}(\omega)$, and subsequently the absorption coefficient following equation 2.98, a fundamental quantity to reproduce absorption spectra.

2.6.3 From DFT and Linear Response Theory to the Computation of Absorption Spectra

Within the linear response framework introduced above, the quantities derived from Density Functional Theory can be systematically connected to experimentally measurable optical spectra. In particular, the central observable in optical spectroscopy is the frequency-dependent absorption coefficient $\alpha(\omega)$, as it is the typical quantity plotted in absorption spectra. $\alpha(\omega)$ is directly related to the complex dielectric function $\varepsilon(\omega)$ as reported by equation 2.98, whose imaginary part encodes the dissipative processes associated with photon absorption.

The imaginary part of the complex dielectric function $\varepsilon(\omega)$, the dielectric tensor $\varepsilon_{\alpha\beta}^{(2)}(\omega)$ is averaged over the volume Ω as in equation 2.99, since optical experiments probe a macroscopic response, whereas first-principles calculations yield a microscopic, spatially varying dielectric function. This averaged $\varepsilon(\omega)$ becomes then a macroscopic quantity that can be directly related to experimentally measurable optical properties, such as the absorption coefficient $\alpha(\omega)$, which, to recall, quantifies the exponential attenuation of light intensity in the material.

Equation 2.97 can be rewritten by considering that the complex refractive index $\tilde{n}(\omega) = n(\omega) + i\kappa(\omega)$ is related to the absorption coefficient, through the formula $\alpha(\omega) = \frac{2\omega}{c} \kappa(\omega)$. Thus, the complex dielectric function $\varepsilon(\omega)$ is related to $\tilde{n}(\omega)$ *via*:

$$\varepsilon(\omega) = \tilde{n}^2(\omega), \quad (2.106)$$

and the imaginary part can be expressed as:

$$\varepsilon^{(2)}(\omega) = 2n(\omega)\kappa(\omega). \quad (2.107)$$

where $\kappa(\omega)$ is the imaginary part of $\tilde{n}(\omega)$. Combining all these expressions yields a linear relationship between the absorption coefficient and the spatially averaged imaginary dielectric function:

$$\alpha(\omega) = \frac{\omega}{c} \frac{\bar{\varepsilon}_{\alpha\beta}^{(2)}(\omega)}{n(\omega)}. \quad (2.108)$$

The proportionality factors involving ω and $n(\omega)$ are commonly referred to as weighting factors. Therefore, plotting $\bar{\varepsilon}_{\alpha\beta}^{(2)}(\omega)$ versus photon energy provides a direct representation of the optical absorption spectrum. In anisotropic systems, different Cartesian components (α, β) can give

different absorption features, but a spatial or orientational average is usually performed to produce a scalar absorption spectrum comparable to experimental measurements, as explained above.

In addition, also equation 2.104 establishes the direct link between the dielectric response obtained from DFT and the experimentally observed absorption profile. As stated in section 2.6.1, the spectral features of $\alpha(\omega)$ are governed by two fundamental ingredients: the energy differences $E_{c\mathbf{k}} - E_{v\mathbf{k}}$, which determine the position of absorption peaks, and the transition matrix elements $|M_{cv}(\mathbf{k})|^2$, which modulate their intensity. The Dirac delta function ensures that only vertical (i.e., crystal momentum conserving) transitions contribute, reflecting the dipole approximation valid for optical wavelengths.

It is important to emphasize that, within standard Kohn–Sham DFT and the independent particle approximation, electron–hole interactions are neglected. As a consequence, the computed absorption spectra correspond to a picture of non–interacting electron–hole pairs, often referred to as the independent quasiparticle approximation. This typically results in an underestimation of band gaps and the absence of excitonic effects, which in real materials can give rise to bound states below the continuum and significantly reshape the absorption onset. More advanced approaches, such as many–body perturbation theory within the GW approximation combined with the Bethe–Salpeter equation, are required to capture these effects.

Nevertheless, even at the independent particle level, DFT–based linear response calculations provide a powerful and widely used tool to interpret plotted absorption spectra. They allow one to assign specific features in the spectra to interband transitions at particular regions of the Brillouin zone, analyze the polarization dependence of the optical response through the dielectric tensor $\varepsilon_2^{\alpha\beta}(\omega)$, and gain microscopic insight into how the electronic structure of a material governs its interaction with light.

Atomistic Insights into Silica–Biomolecule Interactions for APT sample preparation via MD simulations

This Chapter is primarily inspired by Refs. [115, 116]. In particular, the opening section highlights recent experimental advances in the application of the Atom Probe Tomography (APT) protocol in ref. [82] to novel protein structures, with a focus on the Green Fluorescent Protein (GFP). These recent outcomes were directly produced by our collaborators from the Chalmers University (Gothenburg, Sweden), within the MIMOSA project. Such section establishes the foundational framework for the whole Chapter, motivating the use of MD simulations to investigate the interactions between silica and biomolecules, which constitute the central focus. To recall, details about the Molecular Dynamics algorithm can be found in Chapter 2, Sections 2.1 and 2.2.

3.1 Motivating MD Simulations: Recent Experimental Studies on GFP *via* APT

Recently, GFP was analysed through the whole APT protocol described meticulously in Chapter 1, with the same sample preparation protocol and laser irradiation procedure (Section 1.5). This represents the second proof of concept of the whole procedure. The silica matrix containing the protein samples has $\text{Si}(\text{OH})_4$ as precursor molecule and was prepared *via* the same sol-gel method of Ref. [82]. Lastly, isotope labelling was employed for the GFP protein synthesis, to get an unambiguous identification in the mass spectrum obtained during the APT analysis (the ratio $^{12}\text{C}/^{13}\text{C}$ was 1:9 in the composition of the ^{13}C -labelled GFP). The needle-shaped specimens were prepared through the FIB-SEM procedure illustrated in Fig. 1.14 and the APT analysis was performed with a LEAP 6000 XR (Cameca, Madison WI) instrument equipped with a deep-UV laser (257.5 nm).

Results from the whole APT pipeline applied on GFP are reported in Fig. 3.1. These findings come from a manuscript that is currently under submission phases, cited in the Preface of this thesis. The full structural profile of the needle-shaped specimen is reported in Fig. 3.1

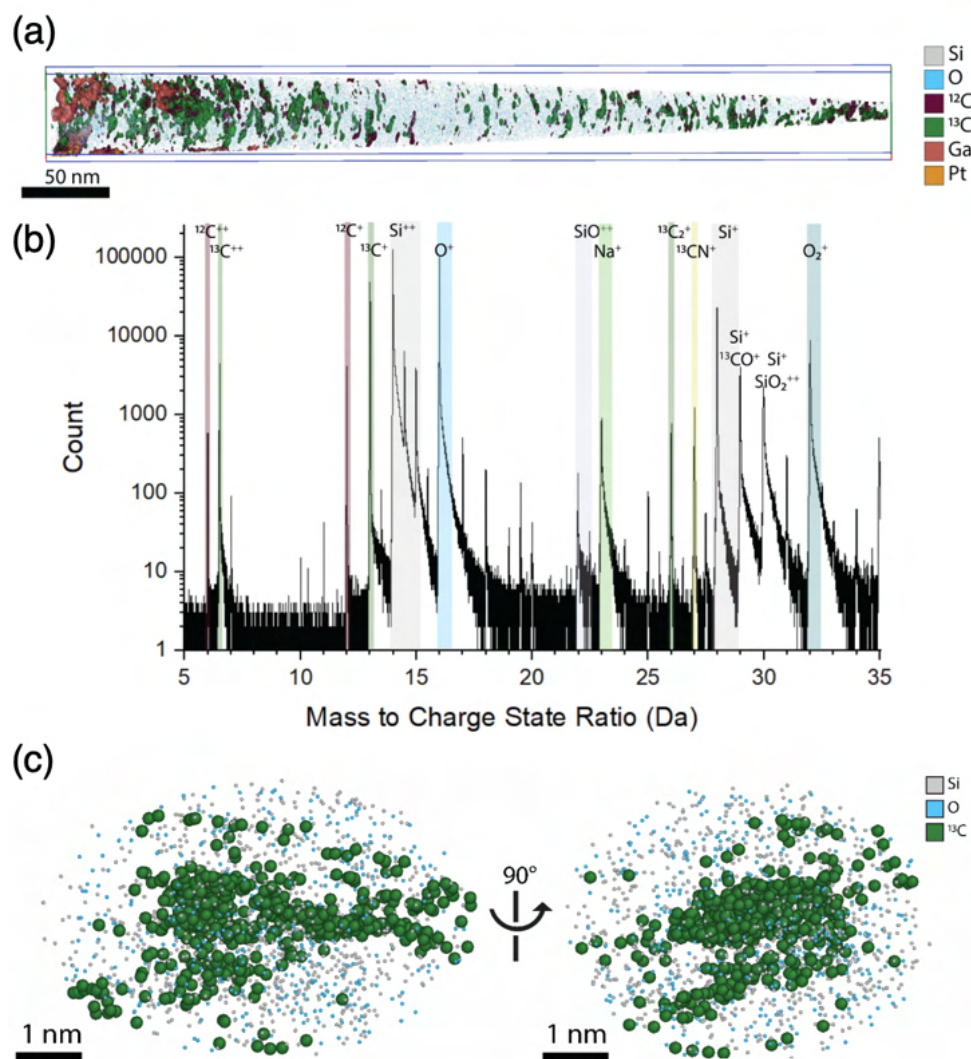


Figure 3.1: (a) Atom map of the reconstructed needle-shaped silica specimen containing ^{13}C -labelled GFP proteins. Silicon atoms are shown as grey dots and oxygen as blue dots. ^{12}C , ^{13}C , Ga, and Pt are shown as isoconcentration surfaces at 2, 20, 5, and 2 atomic % respectively. (b) Segment of the isotope-labeled Time-of-Flight-Secondary-Ion Mass Spectrum (TOF-SIMS) produced during the APT analysis, highlighting the chemical identity of the main peaks referring to GFP and silica. (c) Segment of the atom map in (a) defined by a region of interest (ROI) with the dimensions 4-5-7 nm, isolating a ^{13}C cluster exhibiting size and shape consistent with the known structure of a single GFP molecule. The structure of a GFP molecule can be found in Fig. 3.2.

3.1 Motivating MD Simulations: Recent Experimental Studies on GFP *via* APT85

(a): the ^{13}C atoms (green) were inhomogeneously distributed, frequently forming clusters of approximately 500 atoms, matching the expected 1134 carbon atoms of a single GFP molecule, given a detection efficiency of 52% declared by the instrument. Also few larger clusters were observed, that can be motivated by aggregation phenomena displayed by GFP units during the preparation of the silica matrix, specifically the polymerization step. As discussed in Chapter 1, the role of pH plays a crucial role in the specimen generation phase, and it is thus reasonable to expect that some local pH effects might have led to the occurrence of GFP aggregates within the silica tip. Fig. 3.1 (c) represents the main, core result of this analysis, reporting a ^{13}C cluster exhibiting size and shape consistent with the known structure of single GFP molecule. Overall, this outcome suggests that data can be in principle collected also from small individual proteins like GFP. However, the reconstructed profile lacks specific details, exhibiting a modest resolved pattern. As aforementioned, this whole protocol is far from being completely and fully optimized: errors and artifacts might arise from several steps, the reconstruction algorithm or from the APT instrument itself, as described in Chapter 1, Sections 1.5.3 and 1.5.4. Understanding the origin of these vacancies is essential for improving the reliability and accuracy of the overall protocol, but disentangling their exact sources is a complex task, due to the multiple and interconnected stages of the workflow. For this reason, the present chapter does not aim to provide a comprehensive analysis of all possible error sources. Instead, it focuses primarily on the current sample preparation protocol, with the goal of examining its critical aspects and identifying potential avenues for refinement. There are few studies in the literature regarding the interaction of silica and proteins, that come most from experiments [117, 118, 119, 120, 121]. However, no *in silico* description is present up to now. Thus, a thorough microscopic characterization of the silica adsorption process (and the structural implications thereof) by means of computational techniques is called for. In fact, numerical simulations have been successfully applied to model key steps of the standard APT pipeline. For instance, the occurrence of artifacts through the APT reconstruction phase has been assessed by Molecular Dynamics (MD) and Monte Carlo (MC) methods, while a mixed approach based on MD and *ab initio* calculations has been recently employed to characterize the evaporation mechanism driven by the twofold (laser, static) APT field [122, 123].

3.2 The MD Study

3.2.1 Aim of The Study

In this chapter, classical MD simulations are employed to perform an exhaustive assessment (at the molecular scale) of the adsorption of silica onto diverse biomolecular substrates, specifically a set of small proteins and peptides, displaying diverse structural features. As mentioned in Chapter 1, Sections 1.5.1 and 1.5.3, the silica-embedding APT protocol involves two steps: **early stages** – dissolution of the biomolecule in a solution containing the matrix precursor $\text{Si}(\text{OH})_4$ and some of its short oligomers, as the matrix is polymerizing when proteins are added; **late stages** – the subsequent gradual but stronger condensation of silica into a gel-like/glassy matrix embedding the proteins. This MD study aims at modeling the former process, while the latter will be covered in a small section at the end of this chapter. To recall, the spontaneous encapsulation of biomolecular targets by silica species prior to the laser ablation step constitutes a critical stage in the APT experimental workflow, as it directly influences both sample stability and the fidelity of the reconstructed chemical information. In fact, a primary concern is that the silica matrix must act as an inert scaffold, preserving the native conformation of the protein while reducing thermal fluctuations that could degrade spatial resolution in the following APT analysis. In other words, an excessive interaction with the silica (particularly at high concentrations) could perturb the biomolecule, specifically if its structural features significantly differ from a large and robust IgG antibody. Secondly, water shells around the protein surface must be replaced by silica already in the early stages of the process, as residual H_2O in the final needle-shaped specimen leads to failure during la-APT analyses.

In this context, understanding the physicochemical principles governing the adsorption of silica onto protein surfaces is of paramount importance. To this end, extensive unbiased molecular dynamics (MD) simulations performed on the microsecond timescale provide a detailed atomistic description of the encapsulation process across multiple protein systems and silica compositions.

3.2.2 Biomolecular Systems

Every analysed biomolecular system contains an organic (protein) and inorganic (silica) moiety (see Fig. 3.2). Specifically, the chosen proteins span both folded and intrinsically-disordered motifs, and involve ubiquitin (Ubq), the small ubiquitin-like modifier 1 (SUMO-1), a member of the Enhanced Green Fluorescent Protein (EGFP) family, lysozyme T4 (LT4), a partially folded structure of the amyloid-beta 1-40 ($A\beta_{1-40}$), and a transmembrane helix of the Interleukin-22 receptor alpha-1 (IL22RA1). Ubiquitin is associated with a well-characterized and stable

conformation in solution, and is here employed as a putative benchmark scenario. SUMO-1 is equipped with a ubiquitin-like core and intrinsically disordered carboxy and amino-termini, whose behavior is characterized in terms of their inherent flexibility and motility. The EGFP and LT4 proteins were taken as two almost fully folded proteins, the former characterized by β -sheets organized as a β -barrell, the latter displaying mainly α -helical secondary structures. Contrary, $A\beta_{1-40}$ represents the opposite case, an intrinsically disordered (ID) peptide apart from a short unstable α -helix. Lastly, IL22Ra1 is here assessed as representative of an unstable, purely hydrophobic peptide in solution.

Additionally, the early stages of the APT protocol and the synthesis of the silica matrix were depicted in an effective manner by monitoring the surface affinity of three silica oligomers, i.e., monomer- $\text{Si}(\text{OH})_4$, dimer- $\text{Si}_2\text{O}_7\text{H}_6$, trimer- $\text{Si}_3\text{O}_{10}\text{H}_8$. Since TEOS did not produce reliable silica matrices, as explained in Chapter 1, Section 1.5.2, the present work focuses exclusively on inorganic orthosilicates, excluding TEOS-derived matrices.

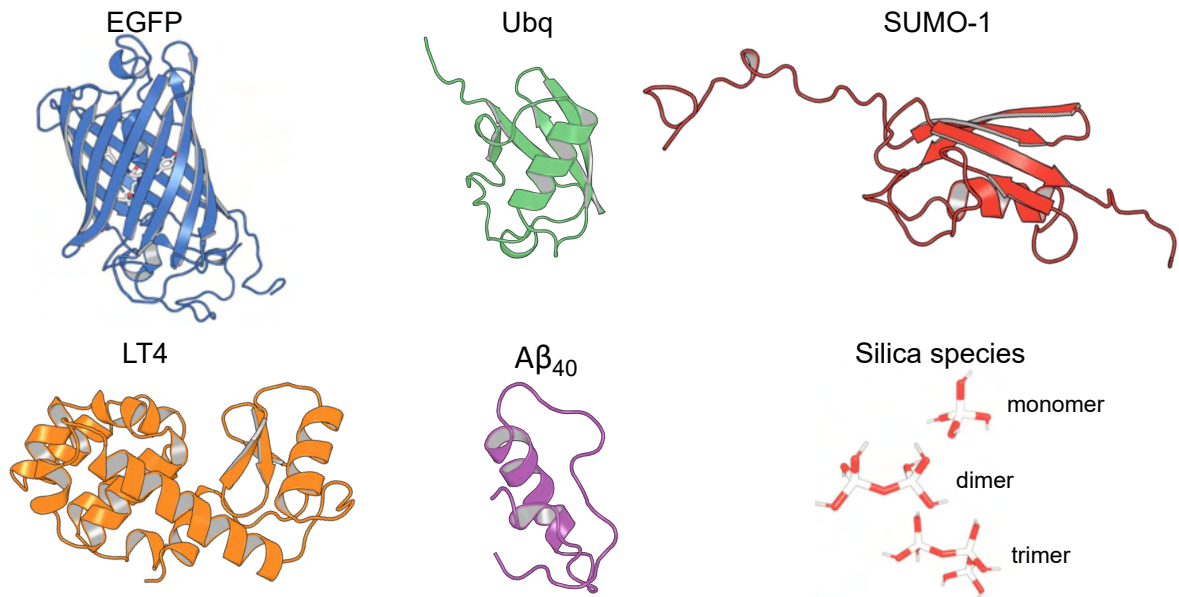


Figure 3.2: Structures of the systems characterized in this work (ILRA22 not reported here) – namely, EGFP (blue), ubiquitin (green), SUMO-1 (red), lysozyme T4 (orange) and $A\beta_{1-40}$ (violet) - and of the silica precursors (i.e., $\text{Si}(\text{OH})_4$, $\text{Si}_2\text{O}_7\text{H}_6$ and $\text{Si}_3\text{O}_{10}\text{H}_8$). Adapted from Ref. [116].

Four main simulation scenarios were thus designed (an example setup is shown in Fig. 3.3):

- **Waterglass**, composed of 204 units of $\text{Si}(\text{OH})_4$ molecule, homogeneously distributed in a physiological solution of water, and ions (150 mM sodium chloride- NaCl); the $\text{Si}(\text{OH})_4$ concentration is 360 mM, compatible with the experimental conditions of the early stages

of the APT sample preparation protocol.

- **Control**, i.e., the system in a physiological solution, built to monitor the natural behavior of each biomolecule;
- **Monomeric**, i.e., the system in a physiological solution added with 204 silica monomers; the $\text{Si}(\text{OH})_4$ concentration is 360 mM, compatible with the experimental conditions. Notably, this case reproduces the very first instances of the APT specimen procedure, in which proteins mainly interact with $\text{Si}(\text{OH})_4$.
- **Mixed**, i.e., the system in a physiological solution added with 34 silica monomers, 34 dimers and 34 trimers. Conceptually, this scenario represents a situation in which some of the $\text{Si}(\text{OH})_4$ molecules in the **monomeric** scenario have polymerized into small oligomers, namely dimers and trimers, as it should occur during the early stages of the APT protocol. Therefore, the total number of silicon atoms is still 204, but these have been redistributed this time in the three silica species sharing the same final concentration of 60 mM.

In the section referred to the results of the work (3.4), also other minor biomolecular systems will be introduced, representing specific variants of the main, core four scenarios. Their conceptualization, characteristics and details are described further on in this chapter.

3.3 Simulation Protocol

This Section has the aim to cover the relevant subtitles of the performed MD calculations, spanning from the chosen molecular atomic coordinates, forcefields and simulation parameters, to simulation setups and performed analysis.

3.3.1 Crystal Structures

The crystal structures of all systems characterized in this work - i.e., ubiquitin (Ubq), SUMO-1, EGFP, lysozyme T4, and $A\beta$ 1–40 - were retrieved from the RCSB Protein Data Bank [124] (entries 1UBQ [125], 1A5R [126], 2Y0G [127], 256L [128], 1AML [129] respectively). For 2Y0G, missing residues at the amino- (16) and carboxy-terminus (8) were reconstructed employing the Modeller interface [130] in UCSF Chimera [131]. The EGFP chromophore was modeled after the structure in Breyfogle *et al.* [132]. The 21- residue sequence of the transmembrane helix of IL22RA1 [133], added with short overhangs at both termini, depicting its native chemical environment (full sequence: DRTWTYSFSGAFLFSMGFLVAVLCYLSYRYV), was taken from the UniProt database [134], and the structure of the helix was built with UCSF Chimera (the full IL22Ra1 sequence is found in [133]).

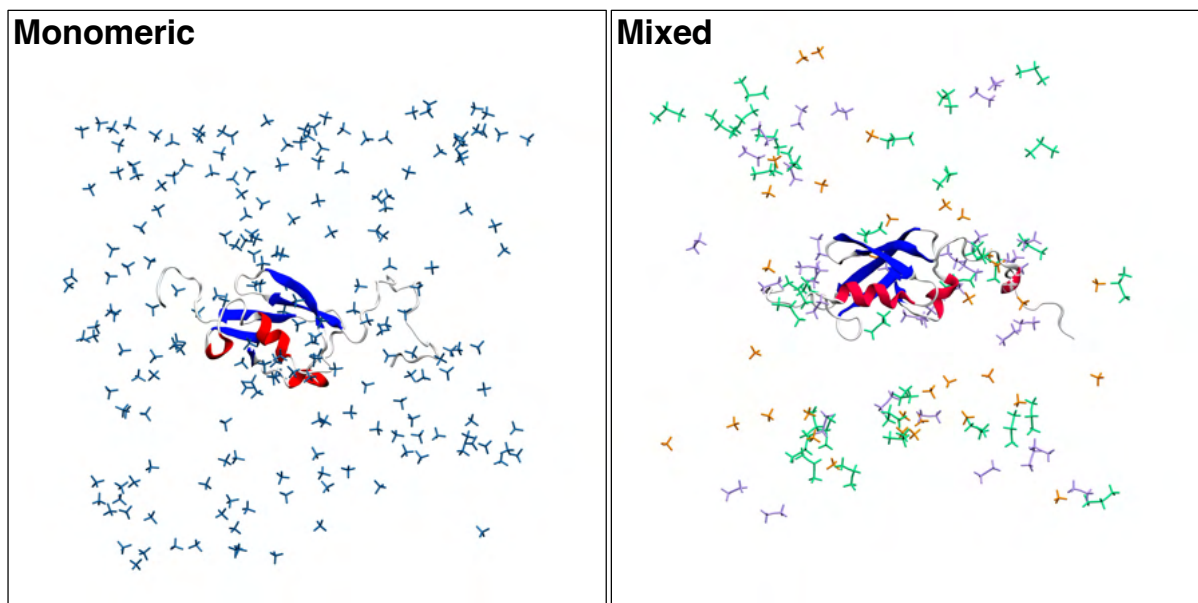


Figure 3.3: Starting configuration of the monomeric and mixed scenarios of SUMO-1 (similar setups have been adopted for every other tested protein): for the latter, silica monomers are colored in orange, dimers in purple and trimers in green. Adapted from Ref. [116]

Lastly, the coordinates of all silica species –i.e., $\text{Si}(\text{OH})_4$ (monomer), $\text{Si}_2\text{O}_7\text{H}_6$ (dimer), $\text{Si}_3\text{O}_{10}\text{H}_8$ (trimer)– were obtained from the PubChem database [135] (CID codes: 14942, 61804, and 11966309 respectively).

3.3.2 Force Fields Selection and Compatibility

Molecular Mechanics techniques are well suited to study complex biomolecular systems, but their reliability depends critically on the availability of accurate and transferable interaction potentials (force fields). A detailed discussion about forcefields and more generally the Molecular Dynamics algorithm is provided in Chapter 2, Sections 2.1 and 2.2). Modeling hybrid systems that combine silicon-based molecules and proteins is inherently challenging, as both components must be described in a way that ensures the resulting potential accurately captures realistic dynamical behavior.

Although silicon is the second most abundant element in the Earth’s crust (28.8% by mass) and naturally occurs in aqueous environments in the form of $\text{Si}(\text{OH})_4$, reaching concentrations of a few ppm in oceans and rivers, until recently no general force field applicable to organosilicon compounds in the liquid phase was available. Previous efforts primarily addressed silica-based solids. For example, force fields were developed to model adsorption processes in zeolites and

other porous silica materials, or to describe crystalline and amorphous silica surfaces [136, 137]. However, these models were optimized for solid-state environments and cannot be directly transferred to molecular systems in the liquid phase. The intramolecular flexibility and intermolecular polarization effects differ substantially between solid silica frameworks and solvated organosilicon molecules.

A significant advancement was achieved by Jorge *et al.* [138], who introduced a versatile and systematically validated force field for organosilicon compounds in the liquid phase. This model, named PolCAFF (Polarization-Consistent Approach Force Field), incorporates approximate treatments of polarization effects while maintaining computational efficiency. Because parametrizing a fully general force field for all organosilicon compounds would be extremely demanding, PolCA focuses on a specific and chemically relevant subset: tetrahedrally substituted organosilicates bearing alkyl ($\equiv\text{SiC}_x\text{H}_y$), alkoxy ($\equiv\text{SiOC}_x\text{H}_y$), and/or silanol ($\equiv\text{SiOH}$) substituents. The alkyl groups are described using a United Atom (UA) approach, in which each CH_x group is treated as a single interaction center (or “bead”) with effective parameters representing the entire group.

This approximation improves computational efficiency but introduces limitations. Molecules such as silane (SiH_4), species containing SiH_3 , SiH_2 , or $\equiv\text{SiH}$ groups, and compounds with aryl substituents are not included in the parametrization. This choice reflects a compromise between generality and practical feasibility, while ensuring compatibility with existing PolCA models for alkanes and alcohols.

Within the framework of MIMOSA, PolCA provides parameters for both TEOS and Si(OH)_4 . TEOS is an alkoxide bearing four OC_2H_5 substituents, whereas Si(OH)_4 contains four silanol groups. However, a methodological issue arises when simulating TEOS together with proteins. TEOS alkyl groups are modeled using the United Atom description, whereas proteins and water are typically described using All Atom force fields. Mixing UA and All Atom representations within the same simulation may introduce inconsistencies in the description of non-bonded and electrostatic interactions. Since TEOS did not produce reliable silica matrices, as explained in Chapter 1, Section 1.5.2, the present work focuses exclusively on orthosilicic acid, Si(OH)_4 , excluding TEOS-derived matrices. In fact, a good aspect of PolCAFF is that the silanol groups of Si(OH)_4 are treated within an All Atom (AA) framework (Si(OH)_4 is an inorganic molecule), ensuring compatibility with standard biomolecular force fields and enabling a more consistent and physically sound description of silica–protein interactions.

Thus, the Amber ff14SB [139], TIP3P [140] and PolCAFF [138] force fields were adopted for the biomolecules, water and silica species respectively, with Joung-Cheatham corrections for sodium chloride ions [141]. The applicability of the Lorentz-Berthelot combination rules for the calculation of Lennard-Jones interactions ensures the formal cross-compatibility between the

Amber ff14SB and PolCA force fields. However, no specific tuning of the interactions has been performed – which would rely on the availability of experimental data and/or on an *a posteriori* validation anyway.

PS: Also, I personally contacted via email the developers of PolCAFF, who ensured that in general the forcefield is compatible with the AMBER methodology & parametrization. Without fine-tuned parameters for the silica-biomolecule interaction, it is fine to accept some source of error in the final behavior of the system, but such deviation from its natural dynamics can be assumed almost insignificant due to the simple chemical structure of the silica species, $\text{Si}(\text{OH})_4$, $\text{Si}_2\text{O}_7\text{H}_6$, $\text{Si}_3\text{O}_{10}\text{H}_8$. Thus, the putative refined parameters (to recall, not available in the literature by any means) should not differ too much from the PolCA ones.

3.3.3 MD setup

Protonation states of biomolecules were checked and verified *via* the H++ online suite [142]. Classical atomistic MD simulations were performed with the GROMACS 2021.5 software [143] in explicit solvent (water) and a 150 mM concentration of monovalent salt (NaCl). All systems/scenarios were subjected to an equilibration protocol as follows: i) 50000 energy minimization steps by the steepest descent algorithm; ii) a first, 1-nanosecond thermalization of the solvent bath in the NVT ensemble at $T = 298$ K employing a Berendsen thermostat [144], associated with a coupling constant $\tau_T = 0.1$ ps; iii) a further, 2-nanosecond volume equilibration in the NPT ensemble at $P = 1$ bar via a Berendsen barostat [144], with $\tau_P = 2.0$ ps and a water compressibility value of $4.5 \cdot 10^{-5} \text{ bar}^{-1}$. In all thermalization stages, all heavy atoms in the solute molecules were restrained by a force constant of $1000 \text{ kJ}/(\text{mol} \cdot \text{nm})$, and a 1 fs integration time step (t_{step}) was applied. The subsequent, unrestrained NPT equilibration of the systems lasted 5 nanoseconds, followed by three 500-nanoseconds MD replicates, amounting to a total $1.5 \mu\text{s}$ per each simulated system. Production simulations were carried out employing a stochastic velocity-rescale thermostat [109] ($\tau_T = 0.1$ ps, $T = 298$ K), and a Parrinello-Rahman barostat [145] ($\tau_P = 2.0$ ps, water compressibility = $4.5 \cdot 10^{-5} \text{ bar}^{-1}$, $P = 1$ bar), and were associated with an integration time step t_{step} of 2 fs. Electrostatic interactions were treated via the PME method [146], with a 1.0-nm cutoff applied to both short-range and Van der Waals interactions. All covalent bonds involving hydrogen atoms were constrained by LINCS [147], and long-range dispersion corrections were applied to both energy and pressure calculations [148].

3.3.4 Details about Analysis and Parameters

The GROMACS 2021.5 analysis toolkits were employed to compute all mean square quantities (RMSF, RMSD) as follows:

$$\text{RMSF}(i) = \sqrt{\frac{1}{N} \sum_{i=1}^N |\mathbf{r}_i(t) - \langle \mathbf{r}_i \rangle|^2} \quad (3.1)$$

$$\text{RMSD}(t) = \sqrt{\frac{1}{N} \sum_{i=1}^N |\mathbf{r}_i(t) - \mathbf{r}_i^{\text{ref}}|^2} \quad (3.2)$$

where $\mathbf{r}_i$ is the position of the atom i , N the total number of atoms in the system. The contact analyses were carried out by fixing a 6.0 Å threshold distance criterion between heavy atoms, taking into account unique, atom-wise contacts with GROMACS 2021.5. The solvent-accessible surface area (SASA) was calculated via the GROMACS implementation of the algorithm described in Ref. [148], applying a probe radius of 1.4 Å. The Radial Distribution Functions were computed with their relative built-in implementation of GROMACS 2021.5. The MDAnalysis [149] and SciPy [150] Python libraries were employed to extract significant biomolecular conformations and to perform a hierarchical clustering analysis of the $A\beta_{1-40}$ dynamics. For the latter, each frame in the aggregated trajectory of $A\beta_{1-40}$ is assigned a string array, whereby each residue is mapped to a secondary structure element (e.g., α -helices, 3-10 helices, β -sheets, etc.) or disordered motif (i.e., coils and turns): thus, the Hamming distance between these strings is calculated and employed as a similarity metric in a hierarchical clustering analysis based on an average-linkage criterion; more information can be found in Fig. 3.4 here below. Iso-density maps of silica occupancy about the systems surface were generated via the VMD 1.9.4 interface [151], applying a grid resolution of 1 Å. An in-house, adapted version of the PyInteraph toolkit [152, 153], built on MDAnalysis libraries, and also Pyinteraph2 [154] were employed to track the occupancy of interactions. In this framework, electrostatic contacts (salt bridges) are defined between acidic and basic moieties within 4.5 Å; hydrophobic contacts occur between centers of mass of hydrophobic side chains within 5 Å; hydrogen bonds were evaluated applying a cutoff distance of 3.5 Å between the donor and the acceptor moieties, and a tolerance of 25 degrees. The protein angular dispersion (PAD_ω) [155], used to characterize the plasticity of residues belonging to the proteins, is defined as:

$$\text{PAD}_\omega = \frac{180}{\pi} \cos^{-1} \left(\frac{1 - CS_\omega}{1 + CS_\omega} \right) \quad (3.3)$$

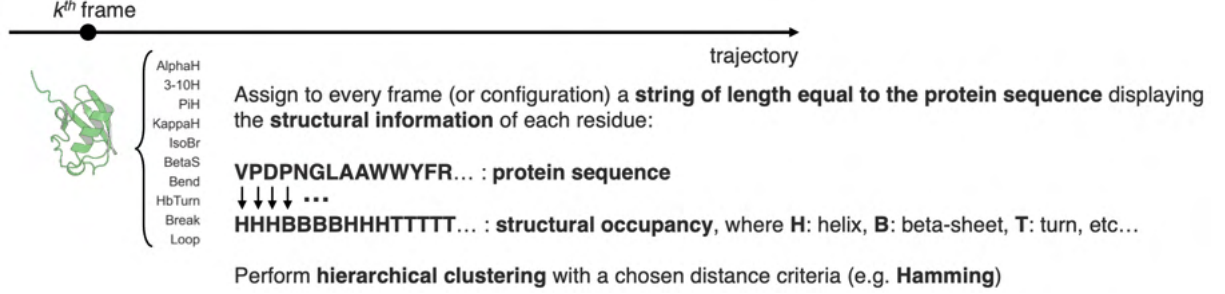


Figure 3.4: Each frame in the aggregated trajectory of $A\beta_{1-40}$ is encoded as a string, wherein each residue is assigned to a secondary structure class (e.g., α -helix, 3_{10} -helix, β -sheet) or to a disordered state (i.e., coil or turn). The similarity between frames is then quantified via the Hamming distance between the corresponding strings, which is subsequently used as the metric in a hierarchical clustering procedure based on an average-linkage criterion.

where CS_{ω} is the circular spread of the angle $\omega = \phi + \psi$:

$$CS_{\omega} = \frac{1 - R_{\omega}^{(2)2}}{2R_{\omega}^{(1)2}} \quad (3.4)$$

Here, $R_{\omega}^{(k)}$ are the k -th order sample moments:

$$R_{\omega}^{(k)} = \frac{1}{N} \left| \sum_{j=1}^N e^{ik\omega_j} \right|, \quad k = 1, 2 \quad (3.5)$$

where the summation runs over N unit vectors sampled from the MD trajectory. PAD_{ω} ranges from 0° (no angular dispersion) to 180° (fully random distribution), providing a quantitative measure of backbone flexibility, representing an alternative to RMSF.

3.4 Results of MD Simulations

This section reports sequentially every result from MD simulations aimed at modeling the early stages of the APT sample preparation protocol. First, outcomes of the Waterglass system are described, in order to get a good assessment of the silica behavior in water. Secondly, the focus shifts toward the protein-silica systems, specifically on a comparison among the Control, Monomeric and Mixed cases. To recall, three 500-nanosecond replicates were generated, amounting for a total simulation time of $1.5 \mu s$, per each simulated system. Details about the molecular systems created can be found in Sections 3.2.1 and 3.2.2.

3.4.1 The Waterglass System: Assessment of the Silica Behavior in Water

The PolCA parameters for $\text{Si}(\text{OH})_4$ are confirmed by DFT calculations

As stated in Section 3.3.2, the utilized forcefield for silica is PolCA. Simulating hybrid systems composed of silicon-derived molecules and proteins presents significant challenges, as the two entities must be modeled so that the whole, final potential model is reliably reproducing a real dynamics. Therefore, we decided to perform standard *ab initio* calculations in order to test the goodness of the principal parameters of the PolCA description of $\text{Si}(\text{OH})_4$.

The equilibrium coordinates of $\text{Si}(\text{OH})_4$ were retrieved by DFT calculations on the Quantum ESPRESSO package [156, 157], employing a plane wave expansion of the electronic wave functions and charge density. The exchange and correlation functional was treated with the generalized gradient approximation, as formulated by Perdew, Burke, and Ernzerhof [158], associated with the optimized norm-conserving Vanderbilt pseudopotentials [159]. Cutoffs of 100 and 400 Ry were chosen for the kinetic energy and the charge density, respectively, to achieve a convergence threshold on the total energy of the system of about 1 meV/atom. A vacuum buffer larger than 15 Å was employed in all calculations, to avoid spurious periodic interactions.

Fig. 3.5 (a) clearly shows that PolCA provides a proper description of bonding and electrostatic terms of the $\text{Si}(\text{OH})_4$ molecule. Moreover, one detail arising from DFT simulations is that the total sum of the calculated partial charges is not equal to that of the formal valence values of each atom. In fact, DFT simulations compute the total charge for one atom by separating the relative values in its orbitals (e.g. s orbitals, π orbitals, ...). However, in many cases, there is a little fraction of charge that does not lay on a specific atom, and cannot be assigned to a specific orbital. This occurrence can be attributed to the phenomenon of delocalization of charge.

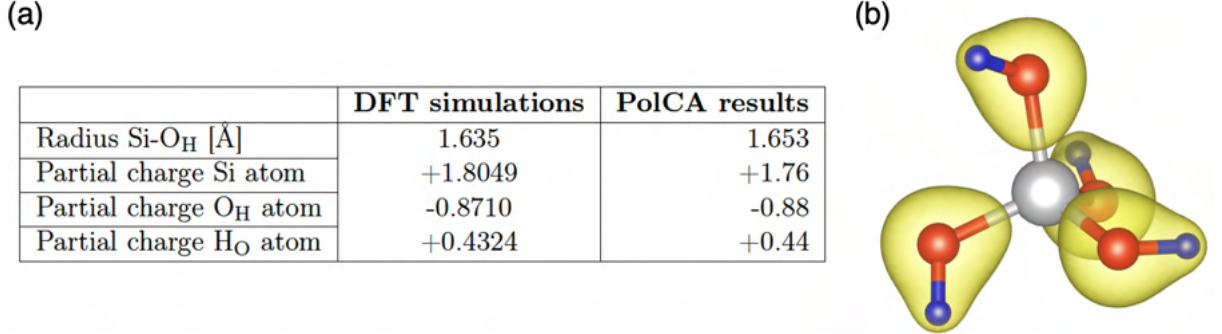


Figure 3.5: DFT calculations on the $\text{Si}(\text{OH})_4$ structure. (a) Comparison between DFT calculations and PolCAFF parameters, i.e. equilibrium distance of Si and O, and atomic partial charges. (b) Representation of the electronic charge density distributed in three-dimensional space (x, y, z) , visualized as the yellow volumetric region in the figure. The isosurface corresponding to a value of $0.1 \text{ e}^-/\text{Bohr}^3$ was selected and rendered.

Equilibrium Dynamics of the Water–Silica Interface: $\text{Si}(\text{OH})_4$ has a low tendency to aggregate

To establish a robust and physically meaningful reference for subsequent simulations involving biomolecular systems, a first in-depth characterization of the equilibrium behavior of a model involving water and silica was performed. This preliminary analysis serves as a critical baseline, allowing us to disentangle intrinsic solvent-mediated effects from protein-specific interactions in later stages of the study. In particular, the focus here is on elucidating the dynamical properties of orthosilicic acid, $\text{Si}(\text{OH})_4$, in an aqueous environment, whether $\text{Si}(\text{OH})_4$ undergoes self-association or not in the probed timescales (nanoseconds). The system was constructed to mimic conditions relevant to atom probe tomography (APT) experiments, where relatively high silica concentrations are often encountered. To this end, the simulation box was prepared with an orthosilicic acid concentration of approximately 360 mM.

The radial distribution function (RDF), $g(r)$, was computed between silicon atoms belonging to distinct orthosilicic acid molecules. The RDF provides a statistical measure of spatial correlations and is defined as:

$$g(r) = \frac{1}{4\pi r^2 \rho} \left\langle \sum_{i \neq j} \delta(r - r_{ij}) \right\rangle, \quad (3.6)$$

where ρ is the number density, r_{ij} is the distance between atoms i and j , and angular brackets denote ensemble averaging. Fig. 3.6 (a) reports the resulting RDF profile between Silicon atoms. $g_{\text{Si},\text{Si}}(r)$ exhibits a smooth and weakly structured behavior, approaching unity at large

distances, which is indicative of a distribution of $\text{Si}(\text{OH})_4$ molecules that is predominantly homogeneous. Importantly, no pronounced peaks are observed at short distances that would signal the formation of stable silica clusters. This observation is further corroborated by hydrogen-bond analysis: the fraction of trajectory frames in which two orthosilicic acid molecules are connected *via* hydrogen bonds remains below 1%, suggesting that intermolecular interactions are both transient and rare (not shown).

Despite the high concentration employed, the system does not exhibit significant silica self-aggregation within the simulated timescale. While such concentrations correspond to a saturated regime, this behavior is fully consistent with the known physicochemical properties of silica in aqueous environments. In particular, aggregation is not governed by simple diffusion-limited encounters, but rather by kinetically hindered chemical reactions.

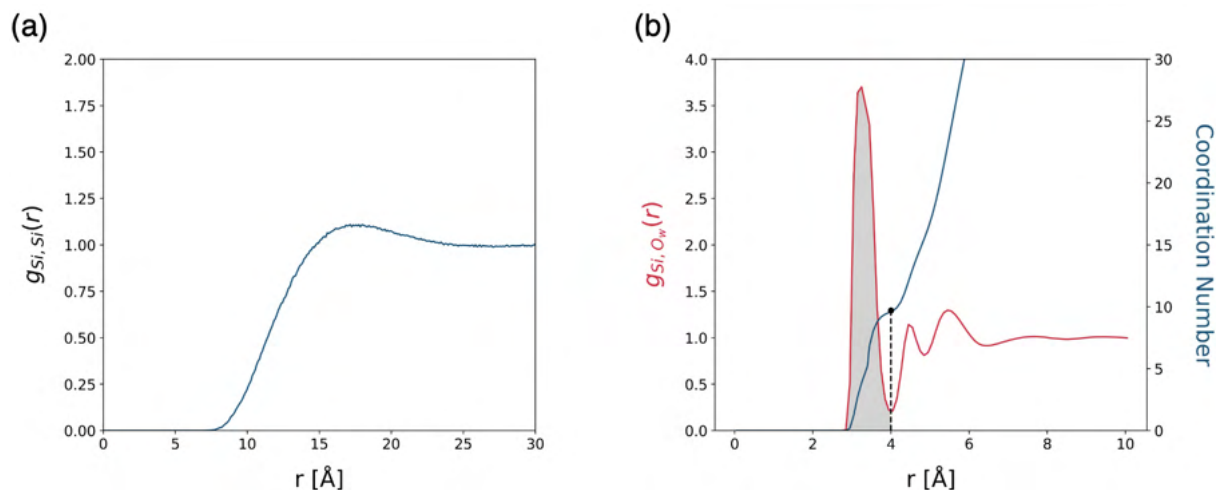


Figure 3.6: Radial Distribution Functions. (a) Radial distribution function $g(r)$ between silicon atoms in the water–silica system, showing a broad and homogeneous spatial distribution of $\text{Si}(\text{OH})_4$ molecules. (b) Radial distribution function of water oxygen atoms around the silicon atoms of $\text{Si}(\text{OH})_4$ (red line), exhibiting a peak at approximately $3.5 \text{ \AA}$. The blue line represents the cumulative integral of $g(r)$, while the shaded area under the peak corresponds to the coordination number of $\text{Si}(\text{OH})_4$, amounting to approximately 10 water molecules per orthosilicic acid molecule. Notably, these RDFs come from a single, total 1.5-microsecond trajectory, composed by the sequential union of the three 500-nanosecond replicates. Adapted from Ref. [115].

The formation of silica oligomers requires a condensation reaction between two orthosilicic acid molecules, involving the formation of a siloxane bridge and the release of a water molecule: $\text{Si}(\text{OH})_4 + \text{Si}(\text{OH})_4 \rightarrow (\text{HO})_3\text{Si}-\text{O}-\text{Si}(\text{OH})_3 + \text{H}_2\text{O}$. This process entails bond rearrangements, including the cleavage of O–H bonds and the formation of Si–O–Si linkages, and is associated with a significant activation free energy barrier. As a result, even when two $\text{Si}(\text{OH})_4$

molecules come into close proximity, the probability of reaction remains quite low, especially in the nanosecond timescale.

In the present framework, the system is described using a classical force field. Consequently, the molecular topology is fixed, and no bond formation or breaking can occur during the simulation. Under these conditions, the condensation reaction described above is intrinsically forbidden, and aggregation beyond transient intermolecular contacts is not possible. The system therefore samples only diffusional motion and weak intermolecular interactions, such as hydrogen bonding. However, even in the presence of a reactive potential, silica condensation is characterized by slow kinetics under ambient conditions. Experimental studies in Refs. [160, 161, 162, 163] report second-order rate constants in the range: $k \sim 10^{-3}\text{--}10^0 \text{ M}^{-1} \text{ s}^{-1}$. At a concentration of $c \approx 0.36 \text{ M}$, the effective reaction rate can be estimated as: $r = kc^2 \sim 10^{-4}\text{--}10^{-1} \text{ s}^{-1}$, corresponding to characteristic timescales of $\tau \sim 10\text{--}10^4 \text{ s}$. The total simulation time in this work is on the order of: $t_{\text{sim}} \sim 500 \text{ ns} = 5 \times 10^{-7} \text{ s}$. A comparison between the simulation time and the characteristic reaction time yields: $\frac{t_{\text{sim}}}{\tau} \sim 10^{-8}\text{--}10^{-11}$, indicating that the simulated dynamics are several orders of magnitude too short to capture even a single condensation event. Therefore, the absence of aggregation is not only expected but unavoidable within the accessible timescale.

This also supports the rationale for employing classical MD as described in Section 3.2.2, i.e. building the Monomeric and Mixed scenarios separately. Because the timescales associated with protein–silica interactions lie in the nanosecond to microsecond range, it is reasonable to assume that each biomolecule experiences silica in a quasi-metastable state of polymerization. Consequently, the effects of ongoing polymerization can be considered negligible.

The Water Hydration Shell of $\text{Si}(\text{OH})_4$

Given that $\text{Si}(\text{OH})_4$ is soluble in water, the organization of water molecules along its volume was characterized by computing the RDF between the silicon atom of $\text{Si}(\text{OH})_4$ and the oxygen atoms of the surrounding H_2O . The resulting distribution $g_{\text{Si},\text{O}_w}(r)$ reveals a well-defined first solvation shell, with a prominent peak centered at approximately $3.5 \text{ \AA}$, as shown by Fig. 3.6 (b). This peak reflects a preferred distance at which water molecules are most likely to be found relative to the silicon center. To quantify the number of H_2O molecules in this hydration layer, the cumulative coordination number, $N(r)$, was defined as:

$$N(r) = 4\pi\rho \int_0^r g(r')r'^2 dr'. \quad (3.7)$$

Integration up to the first minimum of the RDF yields an average of approximately 10 water molecules in the first solvation shell. Moreover, the integration of $g_{\text{Si},\text{Si}}(r)$ up to $3.5 \text{ \AA}$ (peak

maximum) gives 6, a reasonable value also close to literature evidence [160].

In summary, the lack of silica aggregation observed in the present simulations arises from a combination of factors: *i* the intrinsically reaction-limited nature of silica condensation, *ii* the use of a non-reactive force field, *iii* the large discrepancy between simulation and experimental timescales, and *iv* the stabilizing effect of the aqueous environment. These considerations collectively explain why, even under supersaturated conditions, orthosilicic acid remains predominantly monomeric over nanosecond simulation timescales.

Simulation Convergence and Statistical Consistency

To ensure that the observed structural features are not artifacts of insufficient sampling, the convergence of the simulation was assessed by analyzing the time evolution of the RDF. The trajectory was partitioned into contiguous blocks of 25 ns, and RDFs were computed independently for each block. The similarity between these distributions was then quantified using the Jensen–Shannon divergence (JSD), a symmetric and bounded measure of the difference between probability distributions.

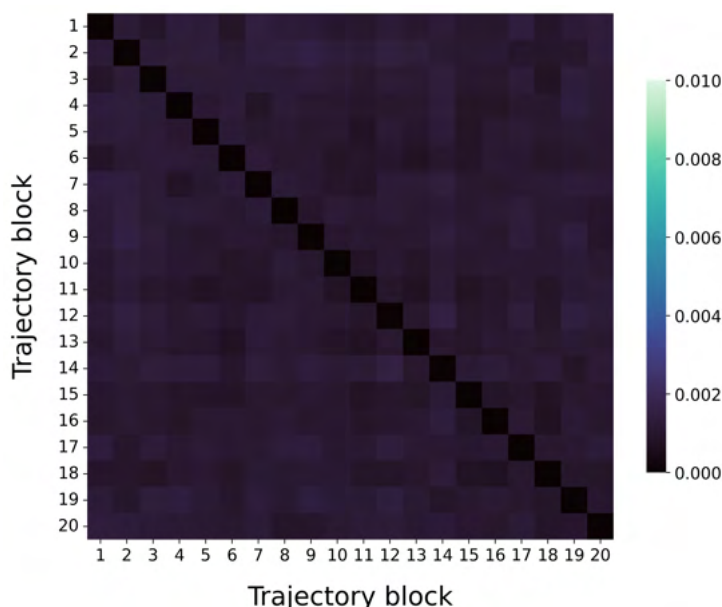


Figure 3.7: Heatmap of the Jensen-Shannon divergence between the radial distribution functions (RDF) associated with each (25-nanosecond) block of the trajectory of the Waterglass system. Notably, these RDFs come from a single, total 1.5-microsecond trajectory, composed by the sequential union of the three 500-nanosecond replicates. Adapted from Ref. [115].

The JSD between two normalized distributions $P(r)$ and $Q(r)$ is defined as:

$$\text{JSD}(P \parallel Q) = \frac{1}{2}\text{KL}(P \parallel M) + \frac{1}{2}\text{KL}(Q \parallel M), \quad (3.8)$$

where $M(r)$ is the average distribution:

$$M(r) = \frac{1}{2} (P(r) + Q(r)), \quad (3.9)$$

and KL denotes the Kullback–Leibler divergence:

$$\text{KL}(P \parallel M) = \sum_r P(r) \log \left(\frac{P(r)}{M(r)} \right). \quad (3.10)$$

with r the inter-atomic distance. Here, the summation runs over discretized distance bins with a resolution of 0.133 Å. Low values of the JSD indicate a high degree of similarity between distributions, and thus convergence of the sampled dynamical properties. This analysis reveals consistently low JSD values across all block pairs, indicating that the RDFs –and by extension, the underlying structural organization of the system– are stable over time. This confirms that the simulation has reached equilibrium and that the observed features of the Waterglass are statistically robust.

Final Remarks

Overall, the present results provide a detailed molecular picture of $\text{Si}(\text{OH})_4$ in water. The absence of significant silica aggregation, combined with the presence of a well-defined yet highly dynamic hydration shell, suggests that H_2O plays a dominant role in stabilizing the orthosilicic acid in solution. The mixture is characterized not by rigid structuring, but rather by a fluctuating network of low-occupancy hydrogen bonds among $\text{Si}(\text{OH})_4$ units that continuously reorganizes on nanosecond timescales. This has important implications for subsequent studies involving biomolecules. In particular, it could suggest that any protein–monomeric $\text{Si}(\text{OH})_4$ interactions observed in later simulations are unlikely to be driven by pre-existing silica clusters, but instead arise from direct interactions with monomeric $\text{Si}(\text{OH})_4$ and its hydration shell. As such, the Waterglass system constitutes a well-defined and physically meaningful baseline for interpreting more complex phenomena.

3.4.2 The Monomeric and Mixed Scenarios: Adsorption Patterns, Structural Determinants, and Domain-Specific Selectivity

This section contains results referred to the following protein structures: Ubiquitin, SUMO-1, EFGP, LT4, $A\beta_{1-40}$. Some evaluations were performed on every protein structure, while others were very specific for a given biomolecule, according to its structural behavior and pattern. The first set of analyses focuses on the silica aggregation and adsorption trend on the protein structures. A subsequent second part is centered on the correlation between adsorption and protein dynamical features. A last, final part is devoted to the analysis of the hydrophobic peptide IL22RA1.

Global Encapsulation Efficiency and Solvent Accessibility Reduction

A first quantitative descriptor of the encapsulation efficiency is provided by the solvent-accessible surface area (SASA). Across all investigated systems (monomeric, mixed) and proteins, a substantial reduction in SASA is observed upon equilibration. Specifically, the silica coating leads to an approximately 45–50% decrease in the exposed protein surface area (see Fig. 3.8), indicating that silica adsorption is both robust and largely independent of the precise composition of the solution. This observation suggests that the driving forces for silica deposition are sufficiently general to operate across different molecular environments, and that even simple silicic acid monomers are capable of contributing to the shielding of the protein surface. However, while the overall degree of encapsulation appears comparable between monomeric and mixed scenarios, a more detailed structural analysis reveals profound differences in the microscopic organization and dynamics of the adsorbed silica layer, that SASA cannot capture.

Spatial Organization of Silica Species: Monomeric versus Oligomeric $\text{Si}(\text{OH})_4$

To characterize the spatial distribution of silica molecules relative to the protein surface, RDFs were computed for all systems (Fig. 3.9). In GROMACS 2021.5, the RDF protein–silica is computed by treating each as a group of atoms and evaluating all pairwise distances between atoms of the two groups, then averaging these distances into a distribution histogram normalized by the bulk density. Specifically, RDFs in the monomeric scenario were computed between the Silicon atoms of $\text{Si}(\text{OH})_4$ and protein residues. In the mixed case, three separate RDFs were evaluated between the Silicon atoms belonging to each species and the biomolecule. These analyses consistently reveal the formation of a well-defined first adsorption shell located within approximately 4 Å from the protein surface, with a pronounced peak centered around 3.5 Å. This feature indicates a preferential accumulation of silica species at close proximity to the protein, involving surface-exposed residues.

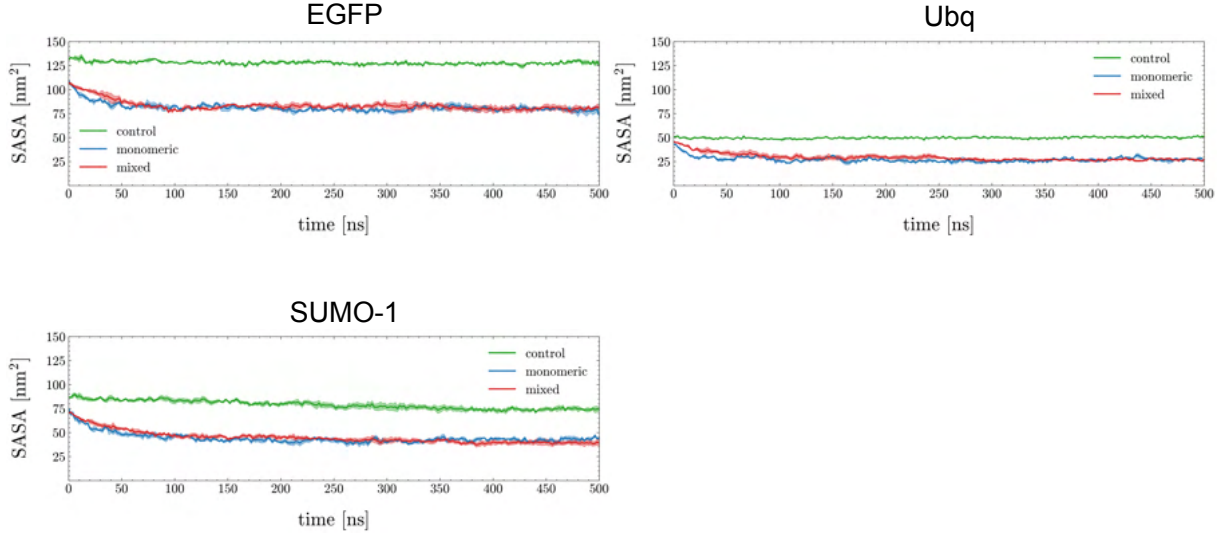


Figure 3.8: Evaluation of the Solvent Accessible Surface Area for three protein structures: EGFP (upper left), Ubiquitin (upper right), SUMO-1 (lower left). Adapted from Ref. [116].

Despite this common pattern, the cumulative RDFs highlight a marked difference between monomeric and mixed systems. In particular, the integrated distribution shows a significantly higher population of silicon atoms near the protein surface that belong to oligomeric species. Dimers $\text{Si}_2\text{O}_7\text{H}_6$ and trimers $\text{Si}_3\text{O}_{10}\text{H}_8$ exhibit a markedly enhanced affinity, as evidenced by their higher probability density within the first few interaction shells. Conversely, silica monomers display a more diffuse distribution, extending more uniformly throughout the simulation volume, as found also from the Waterglass system in Section 3.4.2. This distinction becomes even more evident when considering the radial occupancy up to larger distances. While the majority of dimers and trimers remain localized within approximately 15 Å of the protein surface for all investigated systems (including EGFP, ubiquitin, and SUMO-1), monomers lack such preferential localization and instead populate the bulk solvent more homogeneously, both in the monomeric and mixed scenarios (see Fig. 3.9, right axis, colors blue and orange, respectively). These findings indicate that oligomerization plays a critical role in enhancing the effective interaction between silica and protein substrates.

These observations are further supported by the time evolution of the covering factor (CF), defined as the fraction of silicon atoms located within a specified cutoff distance (4, 6, 8, and 10 Å) from the protein heavy atoms, as shown in Fig. 3.10 (this figure stores data from the EGFP replicates; in the Appendix A the CFs of Ubiquitin and SUMO-1 are reported in Fig. A.3).

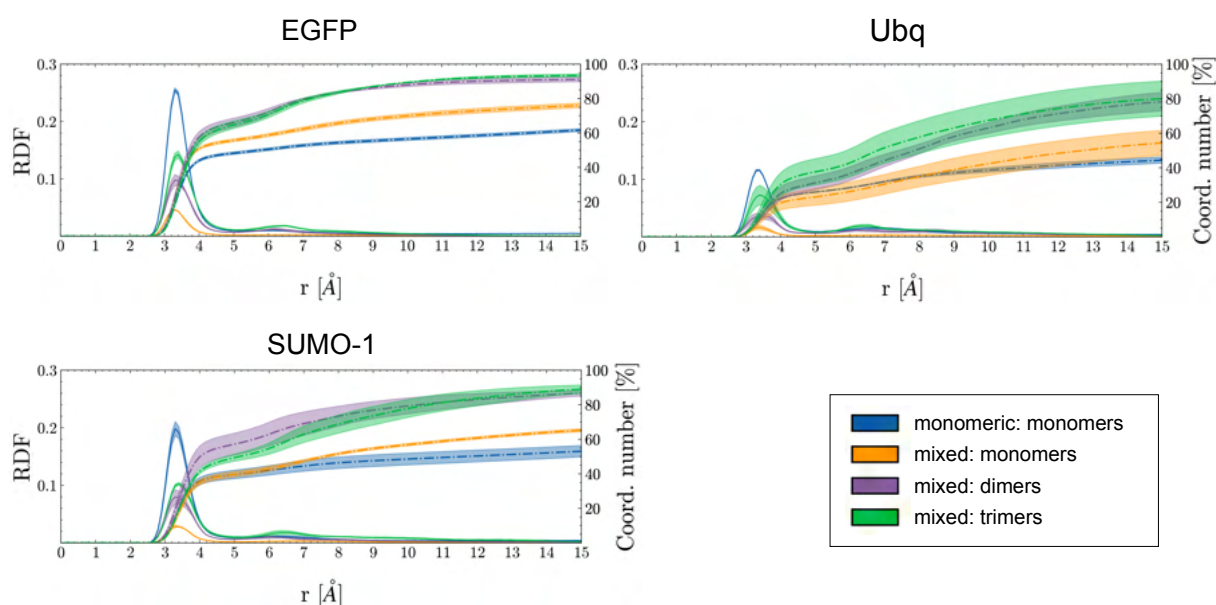


Figure 3.9: **Left axis:** Radial distribution function (RDF) of silicon atoms about the protein surfaces distributed in each silica species $\text{Si}(\text{OH})_4$, $\text{Si}_2\text{O}_7\text{H}_6$, $\text{Si}_3\text{O}_{10}\text{H}_8$, averaged over three 500-nanosecond replicates. Monomeric scenarios are depicted as blue lines, while contributions from silica monomers, dimers and trimers in the mixed scenarios are shown in yellow, purple and green respectively. **Right axis:** Cumulative distributions of the RDFs (the coordination number is reported as a percentage of the total number of molecules per species), highlighting the different adsorption patterns of the silica species on the surface of EGFP, Ubq, and SUMO-1. Adapted from Ref. [116]

Across all systems, silica adsorption proceeds on a characteristic timescale of approximately 100 ns, after which a quasi-steady-state configuration is reached. Importantly, while the overall adsorption pattern is similar in monomeric and mixed scenarios, the contributions of different silica species might differ substantially. Although the CFs were evaluated on the Silicon atom globally (i.e. for the mixed system there is no repartition into each silica species), it is seemingly reasonable to assume that most of the Silicon around the protein structure comes from the oligomeric species. In fact, dimers and trimers should dominate the formation of the stable coating layer, as evinced also by RDF analysis. In contrast, monomeric silica molecules engage in more transient contacts, frequently associating and dissociating from the protein without contributing significantly to long-lived adsorption.

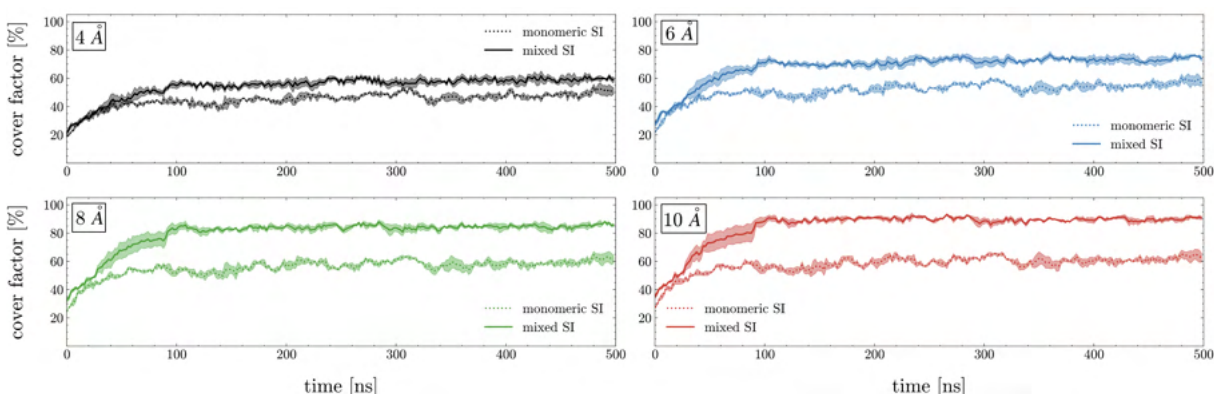


Figure 3.10: Covering factor of EGFP by silica precursors - i.e., the fraction of silicon atoms within 4 Å (black), 6 Å (blue), 8 Å (green), 10 Å (red) of the protein surface (data are averaged over MD replicates). The monomeric and mixed scenarios are shown as dotted and solid lines respectively. Threshold distance criteria are based on the RDFs shown in Fig. 3.9. Results are averaged among the three replicates for each studied system. Adapted from Ref. [116].

To corroborate this, a third analysis about the silica rearrangement on the protein surface was performed. In fact, a key feature emerging from the simulations is the pronounced heterogeneity of the silica coating. Rather than forming a uniform and continuous layer, silica molecules tend to aggregate into localized clusters on the protein surface. This behavior is clearly illustrated by the 3D iso-density maps in Fig. 3.11 and performed in the mixed systems, which show distinct regions of high silica density interspersed with relatively uncoated areas. Such clustering is primarily driven by the intrinsic propensity of oligomeric species to self-associate in aqueous solution, in agreement with previous studies on silicic acid polymerization and aggregation [164, 165]. In fact, the formation of oligomeric aggregates seems to enhance local adsorption by providing multiple interaction sites and stabilizing intermolecular hydrogen-bonding networks.

At first glance, aggregation of oligomers might appear in contrast with previous findings

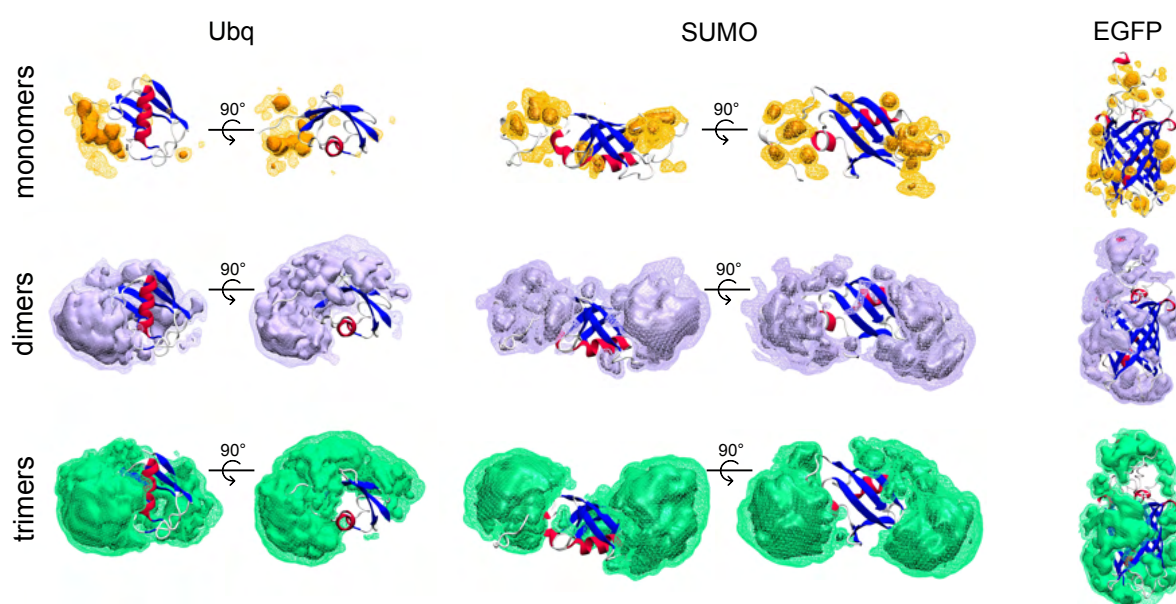


Figure 3.11: 3D density maps of the silica species about the surfaces of EGFP, Ubq and SUMO-1 (expressed in $\text{\AA}^{-3}$) in the mixed scenarios, highlighting the differences in coating affinity and the inhomogeneous nature of the encapsulation process. Iso-density values are depicted as wireframes ($0.10 \text{ \AA}^{-3}$) and darker regions ($0.20 \text{ \AA}^{-3}$). Silica monomers, dimers, and trimers are shown in orange, purple and green respectively. Adapted from Ref. [116].

reported the Waterglass section (3.4.2), where monomeric $\text{Si}(\text{OH})_4$ was shown to remain predominantly dispersed in aqueous solution and not to undergo significant aggregation under comparable conditions. However, the apparent discrepancy can be rationalized by considering the fundamental physicochemical differences between monomeric and oligomeric silica species, as well as the role of the protein surface as a heterogeneous nucleation template.

In bulk solution, monomeric silicic acid is highly solvated and stabilized by surrounding water molecules. The formation of aggregates would require partial desolvation, which is enthalpically unfavorable and not sufficiently compensated by intermolecular interactions between isolated monomers. As a result, $\text{Si}(\text{OH})_4$ remains largely dispersed, consistent with its low tendency to self-condense also at moderately high concentrations. In contrast, silica dimers and trimers exhibit markedly different behavior displaying a higher conformational flexibility, allowing them to adopt geometries that maximize favorable interactions while minimizing steric hindrance. The Si–O–Si bond angle in siloxane bridges is well known for its high intrinsic flexibility, exhibiting a broad angular range (~ 130 – 180°) with a very low energetic penalty (~ 10 kJ/mol) [166].

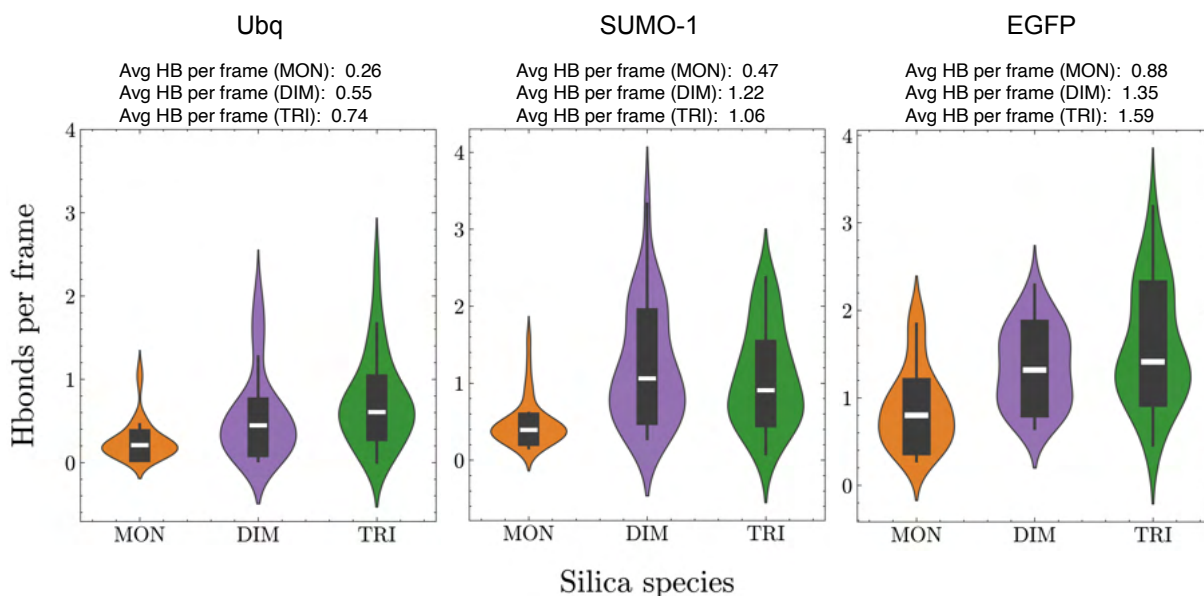


Figure 3.12: Violin plots representing the number of hydrogen bonds per frame (HBpf) between each silica species - i.e., $\text{Si}(\text{OH})_4$ (orange), $\text{Si}_2\text{O}_7\text{H}_6$ (purple), and $\text{Si}_3\text{O}_{10}\text{H}_8$ (green) - and Ubiquitin (left), SUMO-1 (middle), and EGFP (right) in the mixed scenario. The lower and upper edges of the black boxex refer to the first and third quartiles of the HBpf histogram distribution, while the white lines represent the median values. The average values of the HBpf distributions are reported per silica species on top of each plot.. Adapted from Ref. [116].

In addition, the presence of the protein further could shift the balance toward aggregation.

The protein surface probably acts as a heterogeneous scaffold that locally increases the effective concentration of silica species and reduces their translational entropy. Once adsorbed, oligomers experience restricted diffusion, which promotes encounter rates and facilitates cluster formation. In this context, aggregation is not driven solely by bulk solution thermodynamics, but rather by interfacial effects that are absent in homogeneous aqueous environments. In fact, the enhanced affinity of oligomers can be attributed to their increased capacity for hydrogen bonding. Compared to monomeric $\text{Si}(\text{OH})_4$, oligomeric species possess a greater number of hydroxyl groups and exhibit higher conformational versatility, enabling them to establish multiple simultaneous interactions with the protein surface. This hypothesis is supported by the analysis of hydrogen-bonding patterns performed on the mixed scenarios in Fig. 3.12, which shows that oligomers form a higher average number of hydrogen bonds per molecule. The resulting multivalent interactions contribute to stronger and more persistent adsorption, stabilizing the silica coating.

From an energetic standpoint, the clustering of silica oligomers on the protein surface can be interpreted as the result of a balance between enthalpic gains and entropic penalties. The formation of silica–silica and silica–protein hydrogen bonds provides a significant enthalpic stabilization, while the release of structured water molecules from both the protein surface and the hydration shells of the oligomers contributes a favorable entropy gain. This partial desolvation effect is particularly relevant for oligomeric species, whose hydration shells are less tightly bound than those of monomers. Conversely, the loss of translational and rotational entropy upon aggregation is mitigated by the confinement imposed by the protein interface, which effectively pre-organizes the system.

Overall, the absence of aggregation for monomeric $\text{Si}(\text{OH})_4$ in bulk solution and the pronounced clustering of dimers and trimers at the protein interface are not contradictory, but rather reflect two distinct regimes governed by different thermodynamic and structural constraints. While monomers remain stabilized by solvation and lack sufficient interaction strength to self-associate, oligomers can exploit hydrogen bonding affinity and interfacial confinement to form stable, localized clusters on the protein surface.

Correlation Between Silica Adsorption and Protein Structural Features

A central question concerns the origin of the observed heterogeneity in the coating pattern. Specifically, the main focus of this section is to understand whether there is a clear correlation between the silica clusters re-organization on the protein surface and protein structural features,

For instance, in the case of enhanced green fluorescent protein (EGFP) reported in Fig. 3.11, the rigid β -barrel core is partially devoid of silica molecules, whereas the peripheral loop regions display a high density of adsorbed silica. Similarly, for ubiquitin and SUMO-1, the folded cores interact less effectively with silica oligomers than the disordered amino- and carboxy-terminal

regions. These observations might suggest that conformational flexibility and structural disorder could play a key role in mediating silica adsorption.

To quantify these effects, a residue-wise contact analysis was performed, considering the total number of interactions between silica species and protein residues, normalized over simulation time and averaged across independent replicas. The results in Fig. 3.13 confirm that residues belonging to unstructured or flexible regions exhibit significantly higher contact frequencies. In ubiquitin and SUMO-1, the intrinsically disordered termini show particularly strong interactions with dimers and trimers, whereas monomers contribute minimally. A similar pattern is observed for EGFP, where loops and terminal regions dominate the interaction landscape.

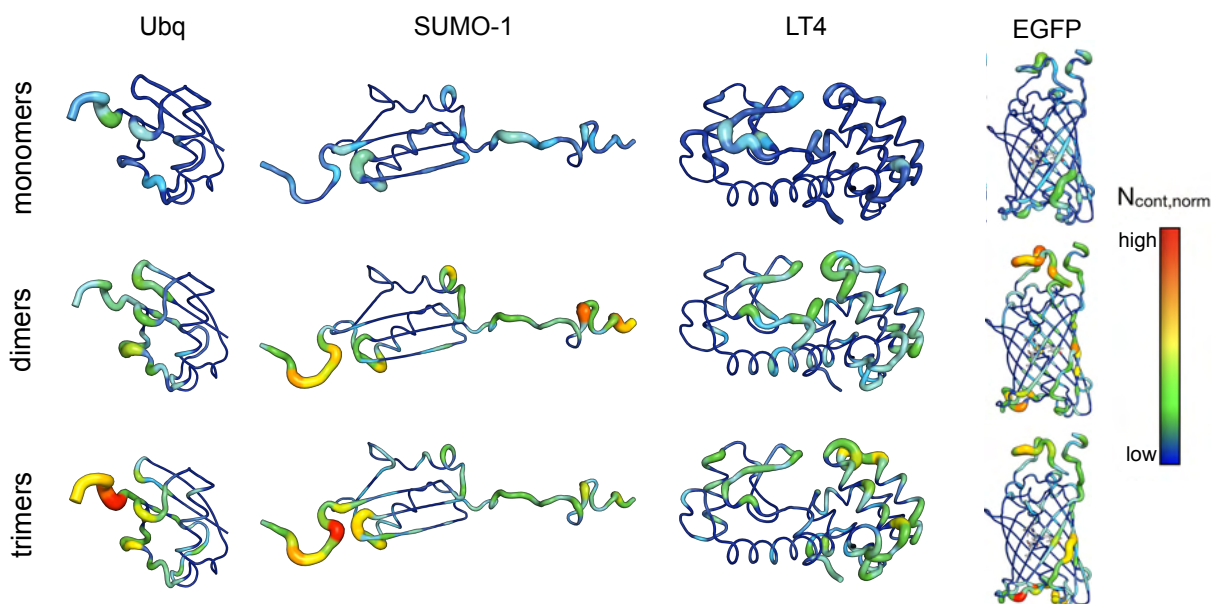


Figure 3.13: Heatmaps of the number of contacts between all silica species and ubiquitin, SUMO-1, lysozyme T4, and EGFP in the mixed scenarios. Contacts are normalized by the number of simulation frames, and averaged across multiple MD replicates. The abundance of contacts is depicted on a color scale ranging from poorly (blue) to densely coated residues (red). Adapted from Ref. [116].

In contrast, lysozyme T4 (LT4), which serves as a model for a fully folded α -helical protein, exhibits a more homogeneous distribution of silica contacts across its surface. Moreover, LT4 displays a reduced overall affinity for silica oligomers compared to the other systems, further supporting the hypothesis that structural rigidity limits adsorption efficiency.

The preferential adsorption of silica onto disordered regions can be rationalized in terms of their physicochemical properties. Intrinsically disordered domains are typically enriched in polar and charged residues, which can act as hydrogen-bond donors and acceptors. This enhances their

ability to interact with the hydroxyl groups present on silicic acid and its oligomers. Additionally, the conformational flexibility of these regions allows them to adapt dynamically to the shape and orientation of the interacting silica species, facilitating the formation of stable binding configurations. In contrast, structured domains impose geometric constraints that limit the number and strength of possible interactions.

Silica encapsulation modulates the structural dynamics of disordered domains

In the previous section it was shown that flexible and intrinsically disordered regions exhibit a markedly different affinity for silica adsorption compared to well-structured domains. At this point, the consequence of this different interaction pattern on the protein dynamics must be investigated. The root-mean-square fluctuation (RMSF) profiles reported in Fig. 3.14 indicate that, in monomeric conditions, the fluctuations of C α atoms remain comparable—or only marginally reduced—relative to control simulations performed in the absence of silica. This observation suggests that isolated Si(OH)₄ molecules exert only a limited influence on the intrinsic flexibility of folded protein structures, as somehow expected.

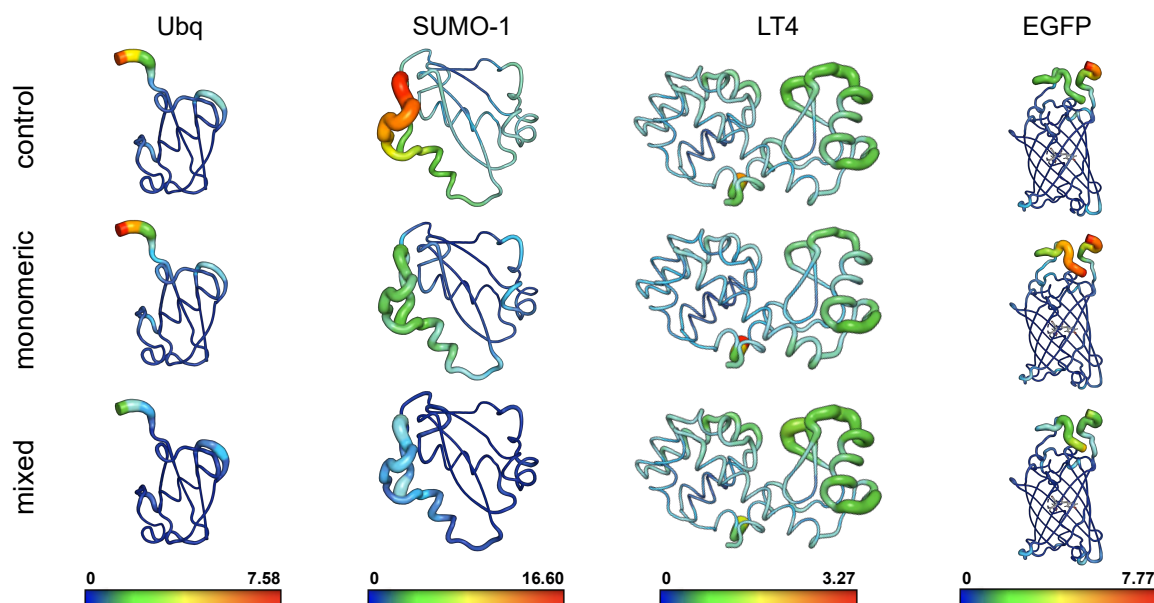


Figure 3.14: RMS fluctuations of the α -carbons of ubiquitin, SUMO-1, lysozyme T4, and EGFP in the control, monomeric, and mixed scenarios respectively, averaged over MD independent replicates. Colors and thicknesses are scaled proportionally to RMSF values, ranging from low (blue/thin) to high fluctuations (red/thick). Adapted from Ref. [116].

In contrast, the presence of silica oligomers in mixed systems leads to a more pronounced

attenuation of structural fluctuations across the investigated biomolecules. This dampening effect is not uniformly distributed but instead preferentially affects structurally labile regions, including intrinsically disordered domains and unstructured termini. Notably, the amino- and carboxy-terminal regions of ubiquitin, SUMO-1, and enhanced green fluorescent protein (EGFP) exhibit a marked reduction in mobility upon silica association, as evidenced by the distinct color gradients observed in Fig. 3.14. These findings indicate that silica adsorption contributes to the early-stage stabilization of flexible protein segments during the formation of the inorganic matrix.

Conversely, well-folded and mechanically robust domains appear largely insensitive to silica encapsulation. This is exemplified by the helix bundle architecture of T4 lysozyme, whose RMSF profiles remain essentially unchanged across control, monomeric, and oligomeric silica conditions. Such behavior underscores the selective nature of silica-protein interactions, whereby structurally rigid motifs retain their native dynamics while flexible regions are preferentially stabilized.

Importantly, the reduction in conformational freedom induced by silica oligomers appears to bias the exploration of the conformational landscape of disordered regions. Rather than uniformly restricting motion, silica encapsulation promotes the selection and persistence of specific locally metastable conformations within an otherwise highly degenerate ensemble. This effect is particularly evident in the case of the amyloid- β fragment $A\beta_{1-40}$, which serves as a prototypical intrinsically disordered peptide.

The structural ensemble of $A\beta_{1-40}$ is predominantly characterized by random coils and transient loop structures, with only occasional formation of an unstable helical segments. However, hierarchical clustering analysis of the aggregated molecular dynamics trajectories (Fig. 3.15) reveals distinct differences between control and silica-containing conditions. While the control scenario displays a general propensity for disorder (lots of middle-ranged colors appear, from light blue to white and orange), the mixed scenario has a more ordered situation. This can be also seen qualitatively by the presence of precise red and blue regions in the mixed heatmap, indicating maximum or minimum occupancy. Notably, a persistent α -helical motif spanning residues 12–24 is observed in the mixed scenario, with a significantly higher occupancy in the presence of silica oligomers. Instead, in the control system, this same region transitions among many helical types, such as α , π and 3-10. More strikingly, residues 30–38 display a marked shift in behavior: whereas they remain largely unstructured in the control simulations, they consistently adopt α -helical conformations in the mixed silica environment.

This observation aligns with previous spectroscopic evidence, including circular dichroism and nuclear magnetic resonance studies, which have demonstrated that oligomeric derivatives of orthosilicic acid can promote local conformational rearrangements in intrinsically disordered peptides, favoring the formation of α -helical structures upon progressive polymerization [167,

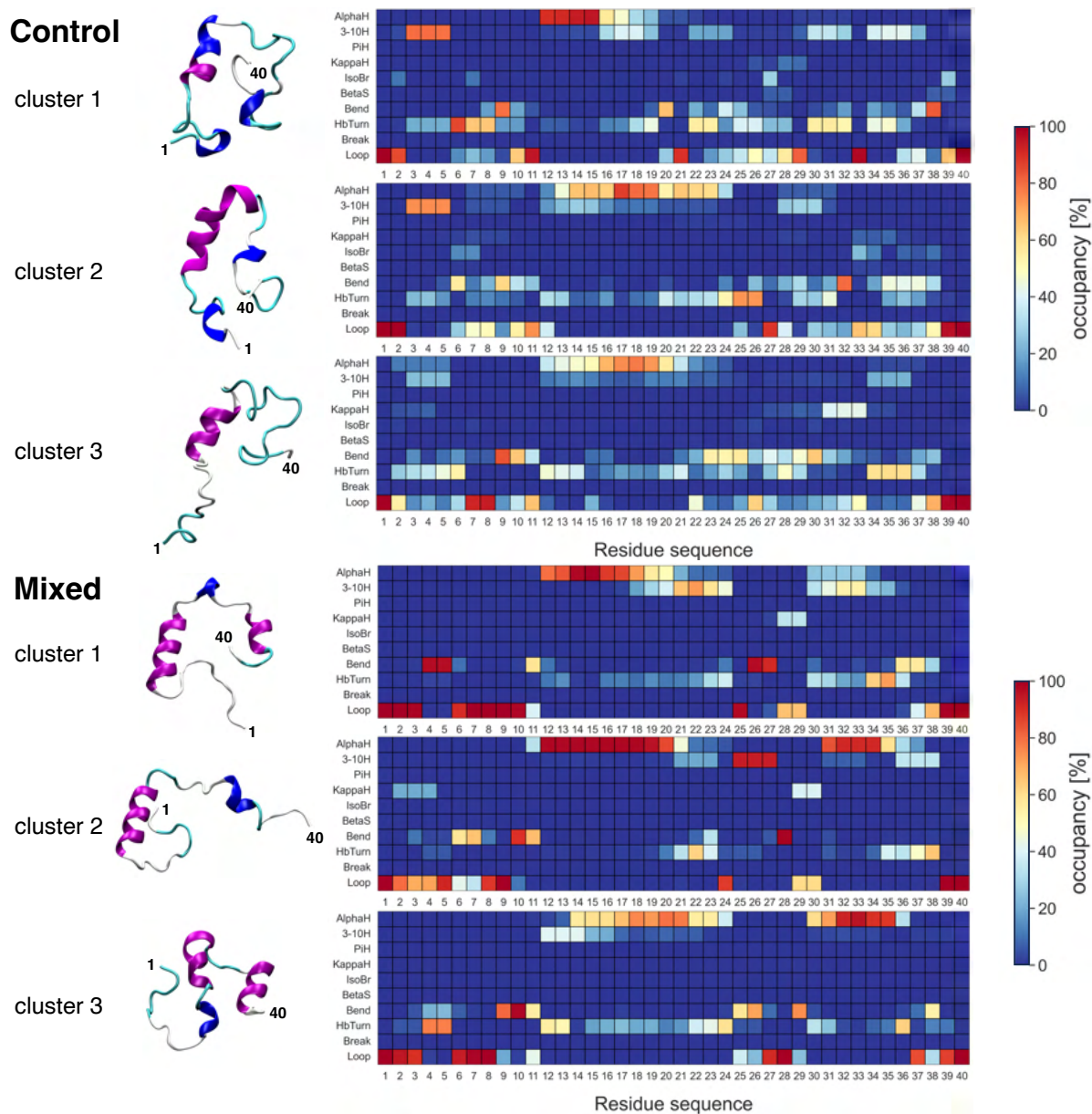


Figure 3.15: Hierarchical clustering analysis of the dynamics of $A\beta_{1-40}$ in the control (top) and mixed (bottom) scenarios, based on the frame-wise evolution of the secondary structure: The heatmaps illustrate the secondary structure propensity of each residue of $A\beta_{1-40}$ within each cluster, categorized into structured (α , 3-10, Pi, and K-helices; isolated bridges, and β -strands) and unstructured (bends, hydrogen-bonded turns, breaks, and loops/coils) domain motifs. For the purpose of clarity, the three most representative clusters of each scenario are shown, along with a representative conformation.

168]. Collectively, these results support a model in which silica encapsulation does not simply immobilize biomolecules but rather reshapes their conformational energy landscape, selectively stabilizing specific structural motifs while preserving overall dynamic behavior.

MD Assessment of the IL22RA1 Peptide System

The last analysis concerns the dynamical behavior of the 31-amino-acids transmembrane helix IL22RA1, here adopted as a prototypical model of a hydrophobic protein segment. It is worth noting that, while such transmembrane helices are natively embedded in lipid bilayers, the present simulations were intentionally carried out in aqueous solution as a simplified and exploratory setup, aimed at probing the intrinsic conformational response of the peptide in relation to the silica rather than reproducing a fully realistic membrane environment.

As expected for a hydrophobic helix solvated in water, the structure rapidly loses its native secondary structure. In both the control and mixed trajectories, the overall helical content decreases by approximately 50% within the first 5 ns, followed by a collapse into a heterogeneous ensemble of compact conformations. Notably, the system does not converge to a well-defined equilibrium structure within the simulation timescale, as confirmed by the root-mean-square deviation (RMSD) profiles reported in Fig. A.1 of the Supplementary Data in the Appendix A. This behavior is consistent with the intrinsic instability of isolated transmembrane helices in polar environments.

To assess whether the unfolding process exhibits distinct conformational features in the two scenarios, the same hierarchical clustering workflow performed in Fig. 3.15 was employed, with the difference that here a full trajectory was created, consisting of the two contiguous control and mixed ones; thus, the analysis was performed on such total trajectory. Given the absence of clear convergence and the rapid interconversion between disordered conformations, a clustering approach based on structural similarity was deemed more appropriate than conventional metrics relying on equilibrium sampling.

The resulting dendrogram (Fig. A.2 – Supplementary Data in the Appendix A) enabled the identification of five significant clusters within the merged trajectory. Fig. 3.15 (a), (b) report both the temporal distribution of these clusters along the simulations and the residue-wise secondary structure occupancy associated with each cluster, respectively. Several trends can be identified from this analysis.

Starting from identical initial conditions, the earliest frames in both control and test (i.e. mixed) scenarios are assigned to a common cluster (cluster 1), as expected. Subsequently, the control trajectory exhibits reversible transitions between clusters characterized by moderate (cluster 3), low (cluster 4), and sparse (cluster 5) helical content, reflecting a highly dynamic and heterogeneous conformational ensemble.

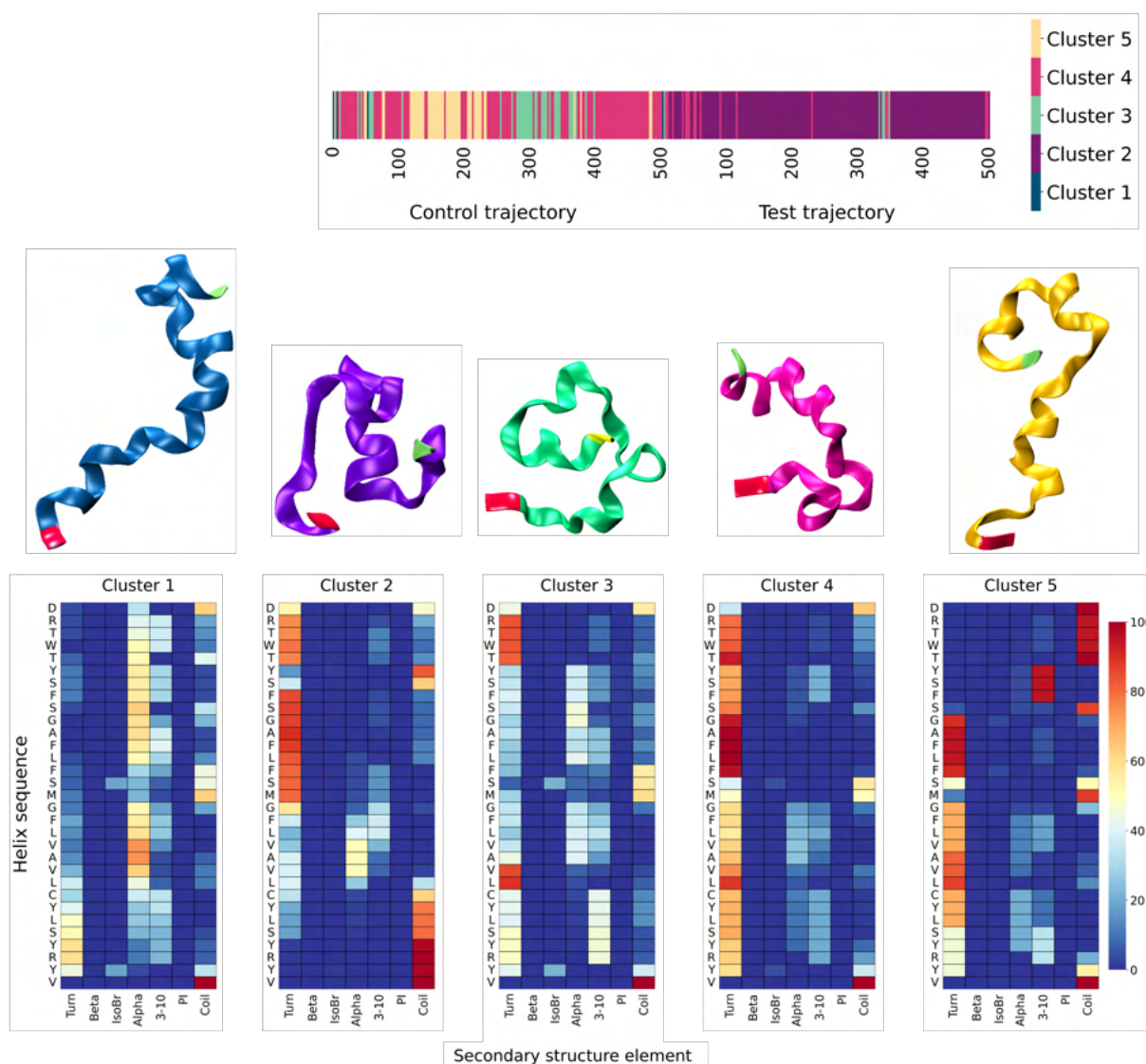


Figure 3.16: **Top:** Timeline distribution of the hierarchical clusters derived from the merged (control, test – here, test refers to mixed) trajectory of the transmembrane helix of IL22RA1 (details in the main text). **Bottom:** Residue-wise occupancy of secondary structure elements and most representative structures associated with each cluster. Adapted from Ref. [115]

In contrast, the test trajectory—i.e., in the presence of silica species—is predominantly associated with a single cluster (cluster 2), which is characterized by the persistence of a small but relatively stable helical core. This residual helix, with an occupancy ranging between 40% and 60%, involves a contiguous stretch of six residues located in the middle portion of the sequence (GLY17–VAL22). This suggests that, despite the overall destabilization of the helix in aque-

ous solution, silica interactions may locally stabilize specific structural motifs,. Such outcome is in analogy with the behavior observed for the intrinsically disordered domains of SUMO-1 and the disordered peptide $A\beta_{1-40}$. In fact, this effect may be interpreted as arising from a viscous buffering exerted by $\text{SSi}(\text{OH})_4$ and its oligomeric derivatives. Such an environment may effectively dampen large-amplitude fluctuations of the peptide backbone, thereby constraining its conformational freedom and promoting the persistence of locally ordered structures.

It is important to emphasize, however, that the observed differences between the control and test scenarios should be interpreted with caution. The limited sampling afforded by 500 ns molecular dynamics simulations, combined with the non-native aqueous environment of the transmembrane helix, may bias the accessible conformational space. Longer simulations and/or alternative setups—such as membrane-mimetic environments—would be required to establish the robustness and generality of these observations regarding the whole IL22RA1 analysis. As stated at the beginning of this section, this evaluation of the IL22RA1 dynamics in relation to water–silica solutions was somehow more of an exploration, rather than a full, detailed investigation.

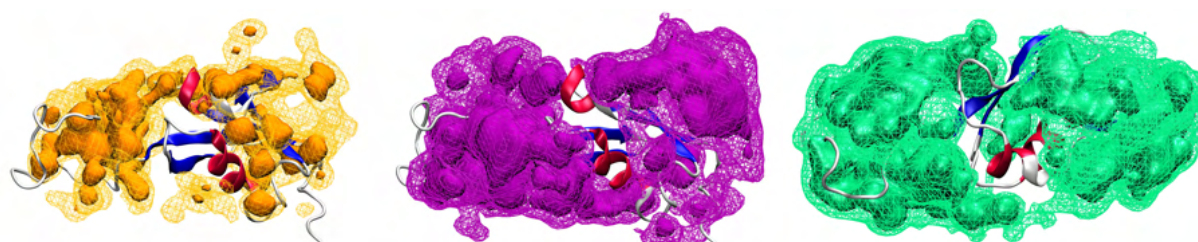
3.4.3 Effects of the Silica Concentration on the Protein Dynamics

This section has the objective to study the effects of the silica concentration on the protein dynamics. Two complementary analyses were carried out. The first concerns an alternative mixed system of only the SUMO-1 protein, designed to probe the effect of silica concentration on the observed protein–silica interactions. The second evaluation sees the creation of three EGFP systems at diverse concentrations of only the orthosilicic acid trimer $\text{Si}_3\text{O}_{10}\text{H}_8$, taken here as proxy of the early stages in the synthesis of the silica matrix.

In the first analysis setup, to verify whether the concentration of silica species (that is, a 1:1:1 ratio between the concentration of silica monomers, dimers, and trimers) affects their adsorption behavior, an alternative mixed scenario was created, where the 204 silicon atoms have been equally split between 23 trimers, 34 dimers and 67 monomers: As such, the concentration of silica monomers is twice the concentration of dimers, and three times the concentration of trimers. Three 500-ns MD replicates of SUMO-1 were performed, taking advantage of its heterogeneous structural framework. Results reported in Fig. 3.17 clearly show that the adsorption behavior of the silica oligomers is seemingly unaffected by their relative ratio, as the results are comparable with the mixed case. However, it is reasonable to assume that, if the concentration of oligomers is too low with respect to the monomeric form $\text{Si}(\text{OH})_4$, the expected behavior could shift closer to the monomeric systems.

The second analysis focuses on simulations at diverse concentrations of a silicate trimer –600, 900, 1200 mM (also here three 500-ns MD replicates were performed). The found isodensity maps in Fig. 3.18 (a), (b) and (c) confirm what found in the mixed scenario, somehow also displaying

SUMO-1 3D density maps: alternative mixed scenario



silica monomers

silica dimers

silica trimers

SUMO-1 RMSF values

control

monomeric

mixed

alternative mixed

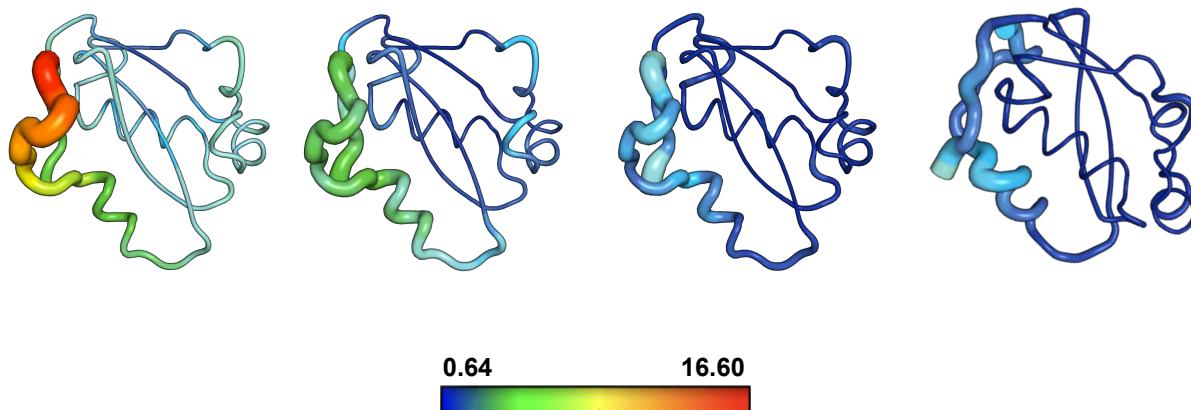


Figure 3.17: **Top** 3D density maps of the silica species about the surface of SUMO-1 (expressed in points·Å⁻³) in the alternative mixed scenario. Iso-density values are depicted as wireframes (0.10 points·Å⁻³) and darker regions (0.20 points·Å⁻³). Silica monomers, dimers, and trimers are shown in orange, purple and green respectively. **Bottom** RMS fluctuations of the α -carbons of SUMO-1 in the control, monomeric, mixed and alternative mixed scenarios respectively, averaged over MD replicates. Colors and thicknesses are scaled proportionally to RMSF values, ranging from low (blue/thin) to high fluctuations (red/thick). Adapted from Ref. [116]

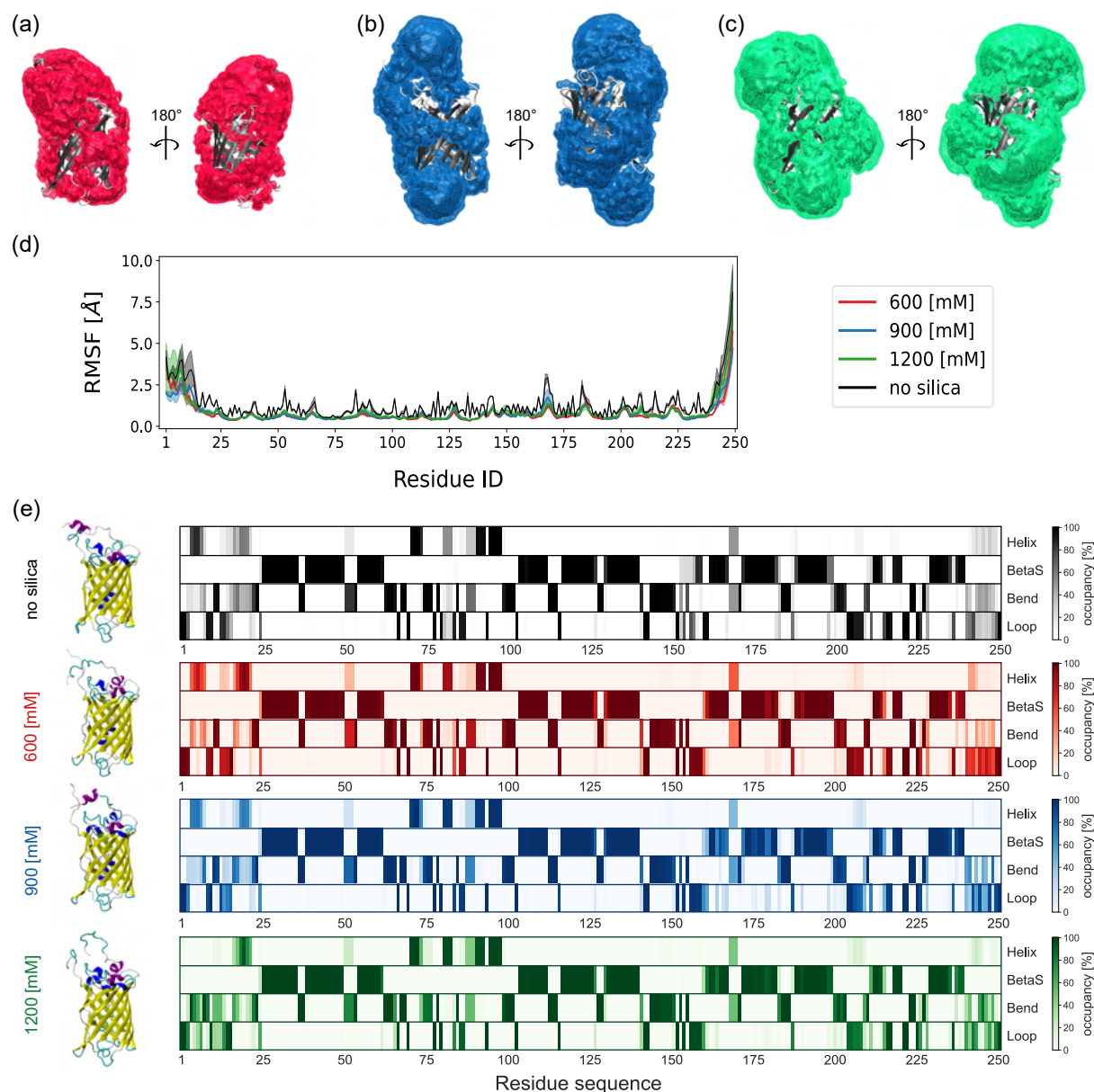


Figure 3.18: Analysis of the EGFP at diverse concentrations of $\text{Si}_3\text{O}_{10}\text{H}_8$. Iso-density maps of the homogeneous adsorption of silicate trimer on the surface of the EGFP, at a (a) 600 mM, (b) 900 mM, and (c) 1200 mM silicate concentration. (d) RMSF of the α -carbon atoms of each residue of the GFP, at diverse concentrations of silicate, with respect to an average structure. (e) Secondary structure propensity of each residue of the EGFP in the four (control, 600 mM, 900 mM, 1200 mM) scenarios, expressed as frame-wise percentage along the MD trajectories. On the left side, the average structures along the three replicates per system are reported.

a major covering effect exerted by the trimeric forms of $\text{Si}(\text{OH})_4$. Interestingly, EGFP retains its secondary and tertiary structure entirely throughout the embedding process, as shown by the secondary structure propensity of each residue in Fig. 3.18 (e). Moreover, the surface adsorption of silica oligomers somewhat dampens the fluctuations of all residues in the sequence (Fig. 3.18 (d)), thereby enhancing the structural stability of the substrate, validating the hypothesis that silica behaves as an inert molecule also at extremely highly saturated regimes.

3.5 Concluding Remarks

A recent advancement in Atom Probe Tomography (APT) has introduced the possibility of encapsulating biomolecules within an amorphous silica matrix, followed by controlled ablation using short laser pulses, leading to a single molecule reconstruction of the target of interest. A crucial requirement for the success of this methodology lies in preserving the native conformation of the biomolecule during the embedding stage, while simultaneously minimizing mechanical stress.

In this study, MD simulations were employed to investigate the adsorption behavior of silica species –specifically the monomer $\text{Si}(\text{OH})_4$, dimer $\text{Si}_2\text{O}_7\text{H}_6$, and trimer $\text{Si}_3\text{O}_{10}\text{H}_8$ of orthosilicic acid– on a variety of biomolecular substrates. These simulations provide insight into the early stages of silica matrix formation within the APT workflow.

The encapsulation of proteins by silica is a complex process governed by the interplay between silica species, protein structural features, and intermolecular interactions. A clear distinction emerges between monomeric and oligomeric silica species: while monomers remain evenly dispersed in solution, oligomers tend to accumulate and organize around the biomolecular surface, replacing water and establishing stable and localized adsorption patterns. This trend persists across different silica concentrations and also at oversaturated regimes.

MD results further indicate that structural features of the substrate play a significant role in silica adsorption. In particular, coiled and intrinsically disordered regions promote stronger interactions with silica species. This behavior can be rationalized by considering the enhanced flexibility and hydrophilic character of these regions, which provide accessible and energetically favorable binding sites. These findings are consistent with previous studies highlighting the importance of amino acid composition and charge distribution in biosilicification processes [169, 170].

Interestingly, the embedding procedure has a limited effect on the conformational stability of well-folded domains, whereas it markedly dampens fluctuations in disordered regions. In some cases, this results in the stabilization of transient secondary structures, such as helical motifs, as observed for intrinsically disordered peptides like $A\beta_{1-40}$. This selective structural influence, also suggested in earlier works [171], points to the potential of APT as a tool for probing the conformational landscape of flexible biomolecular regions at the single-molecule level, although further investigation is warranted.

Overall, this work sheds light on the microscopic mechanisms governing the APT embedding process and supports its application in high-resolution structural characterization of biomolecules. An aspect of relevance is surely the optimization the APT sample preparation protocols. Achieving a controlled and reproducible silica encapsulation is essential to preserve the native-like conformational landscape of biomolecules while ensuring sufficient mechanical

stability and chemical contrast during field evaporation. The demonstrated dependence of protein dynamics on silica speciation and concentration suggests that fine-tuning the conditions of silica polymerization may enable a more precise control over the embedding process, ultimately improving the spatial resolution and reliability of APT measurements on biological specimens. Notably, recent experimental studies by G. Eriksson and co-workers have demonstrated the capability of APT to resolve gold nanoparticles embedded in silica matrices [172]. Furthermore, functionalization of nanoparticle surfaces with proteins prior to encapsulation has been explored as a strategy to improve stability and broaden the applicability of the technique, for instance in studying biomolecule–metal interfaces. Given the rapid progress in APT methodologies, further experimental validation and expanded datasets are expected in the near future.

3.6 Perspectives of the Study: Going Beyond Classical Molecular Dynamics

Until now, the main discussion focused on the application of classical Molecular Dynamics to the reproduction of the early stages of the APT sample preparation protocol. Instead, this section is dedicated to the modeling of the late stages of the protein encapsulation step, specifically trying to address the issue of chemical reactivity and silica polymerization within the Molecular Dynamics framework.

While classical Molecular Dynamics provides valuable insight into the structural and dynamical aspects of the early adsorption process, it inherently lacks the capability to describe bond formation and breaking, essential to capture the transition from dissolved silica species to an amorphous silica matrix. In this context, extending the present study toward reactive simulation approaches represents a natural step. In particular, the explicit modeling of condensation reactions among silica precursors and their coupling with biomolecular surfaces would provide a direct link between the early adsorption stages and the final embedded state observed in APT experiments.

However, addressing such polymerization process is not so easy. First of all, the $\text{Si}(\text{OH})_4$ polymerization might not be compatible with the timescales explored by MD. In fact, spontaneous reactions among silica species does not typically occur on the *nanosecond* timescale at the concentrations used in APT (0.3–0.6 M), but over milliseconds to hours (see Section 3.4.2). Thus, observing bond formation within nanoseconds *in silico* presents a huge challenge, and maybe does not even reproduce what really happens in reality.

Overall, achieving this level of description introduces significant methodological challenges, especially in systems as chemically heterogeneous as protein–silica interfaces. These aspects motivate the exploration of alternative simulation strategies, which are discussed in the following

sections.

3.6.1 Modeling Reactivity: Reactive Molecular Dynamics and Reaxff

ReaxFF is a reactive force field for Molecular Dynamics that enables bond formation and breaking through a distance-dependent bond-order formalism, unlike classical force fields with fixed connectivity [107, 108]. The bond order serves as a bridge between quantum mechanical methods and classical simulations by incorporating QM-trained empirical potentials, thus achieving a balance between accuracy and computational efficiency. A scheme of the ReaxFF formalism is illustrated in Fig. 3.19 (a), while major details can be found in Chapter 2, Section 2.2.2.

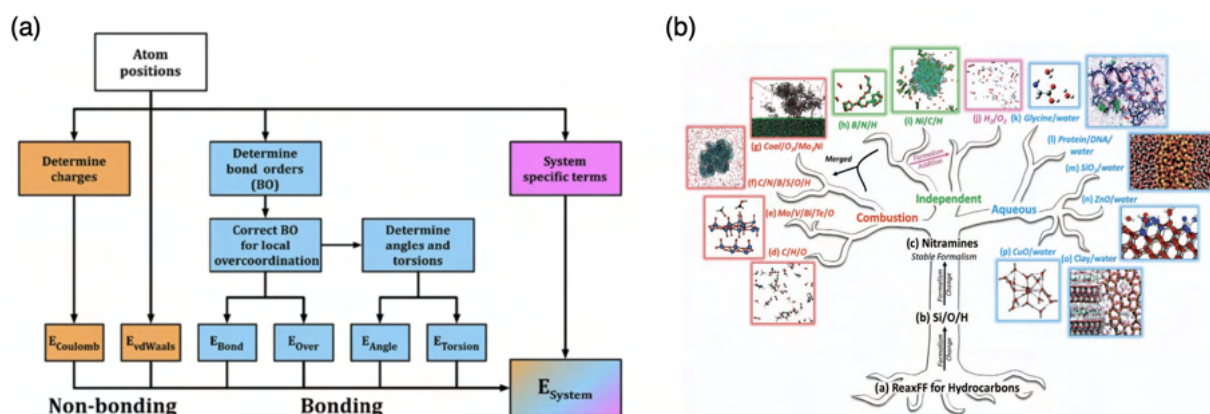


Figure 3.19: **(a)** Energy Evaluation within the ReaxFF framework. Briefly, energy is still divided in non-bonded and bonded contributions. The former are evaluated like in classical force fields, while the latter follow a different treatment. According to bond order measurements, the topology of the system is implicitly guessed; then, the bonded terms are calculated, corrected for local overcoordination and summed up. The final energy of the system results from the addition of the non-bonded part to the bonded one (see Chapter 2, Section 2.2.2 for a more exhaustive treatment of the ReaxFF formalism). **(b)** Organization of the ReaxFF published force fields. Adapted from Ref. [108].

A key characteristic of ReaxFF is the use of a single atom type per element, regardless of its chemical environment. This means that every atom of a given element is parametrized in the same way during a simulation, not caring about its surroundings, e.g. whether it belongs to a biomolecule or a solvent molecule. While this allows for chemical reactivity and flexibility, it reduces the ability to capture subtle environmental differences, introducing at least, approximations or artifacts. Additionally, the reliability of ReaxFF depends on the quality and transferability of its parameterization, which is fitted to specific QM datasets and may not generalize well to different systems.

Finally, the charge equilibration scheme can introduce artifacts such as unphysical charge

transfer, especially in heterogeneous systems. Overall, although ReaxFF is a powerful tool for modeling reactive processes, its results require careful interpretation due to these intrinsic limitations.

The ReaxFF Branches

ReaxFF parameter sets are derived by fitting to quantum mechanical data, including bond dissociation curves, angle distortions, reaction pathways, and condensed-phase properties. Each parameterization is tailored to a specific class of materials, leading to the development of distinct *branches*, see Fig. 3.19 (b).

Examples include silica-focused parameterizations (optimized for Si/O/H systems), hydrocarbon combustion branches (C/H/O), and biomolecular variants (C/N/O/H systems). Within a given branch, parameters can be internally consistent and transferable across related systems, according to the case. However, different branches are *not* inherently compatible. Since all interactions are co-optimized simultaneously, combining parameter sets from different branches generally leads to inconsistencies and artifacts. This is further complicated by the *absence of atom typing*, which prevents the independent tuning of specific chemical environments. Parameter optimization involves a high-dimensional fitting procedure, where competing objectives (e.g., gas-phase accuracy vs condensed-phase behavior) must be balanced. As a result, transferability remains one of the main challenges of ReaxFF.

Silica Polymerization within the ReaxFF Framework

ReaxFF provides a decent silica-focused parameterization, optimized for Si/O/H systems, like for example the Waterglass System. In fact, Dupuis *et al.* [173] investigated the formation of $\text{Si}(\text{OH})_4$ gel-like structures, focusing on the effects of dilution and drying on the resulting matrix. While Reactive Force Fields have not been implemented in GROMACS, LAMMPS [174] was used in their study, the methodology demonstrates that polymerization and vitreous matrix formation are achievable computationally. Water plays a crucial role: drying induces glassy behavior, while hydration softens the gel. Depending on hydration conditions, the atomistic structure of the silica gel can display varying degrees of order or disorder. The employed force field was developed by Fogarty *et al.* [175], which enables reactions between $\text{Si}(\text{OH})_4$ acid molecules, leading to linear and branched $\text{Si}(\text{OH})_4$ chains. This approach allowed them to emulate the vitreous composition of a silica matrix under drying conditions, although no proteins were included. Simulating gel formation remains computationally demanding due to the slow polymerization kinetics of siloxane (Si-O-Si) bonds and the weak long-range order of gels. To overcome these challenges, the study combined Parallel Tempering (PT) to accelerate bond formation with Grand-Canonical Monte Carlo (GCMC) to remove water molecules, effectively

mimicking drying conditions [176, 177, 178]. A scheme of this is reported in Fig. 3.20 (a), that was also reproduced by me with LAMMPS [179] (see Fig. 3.20 (b)).

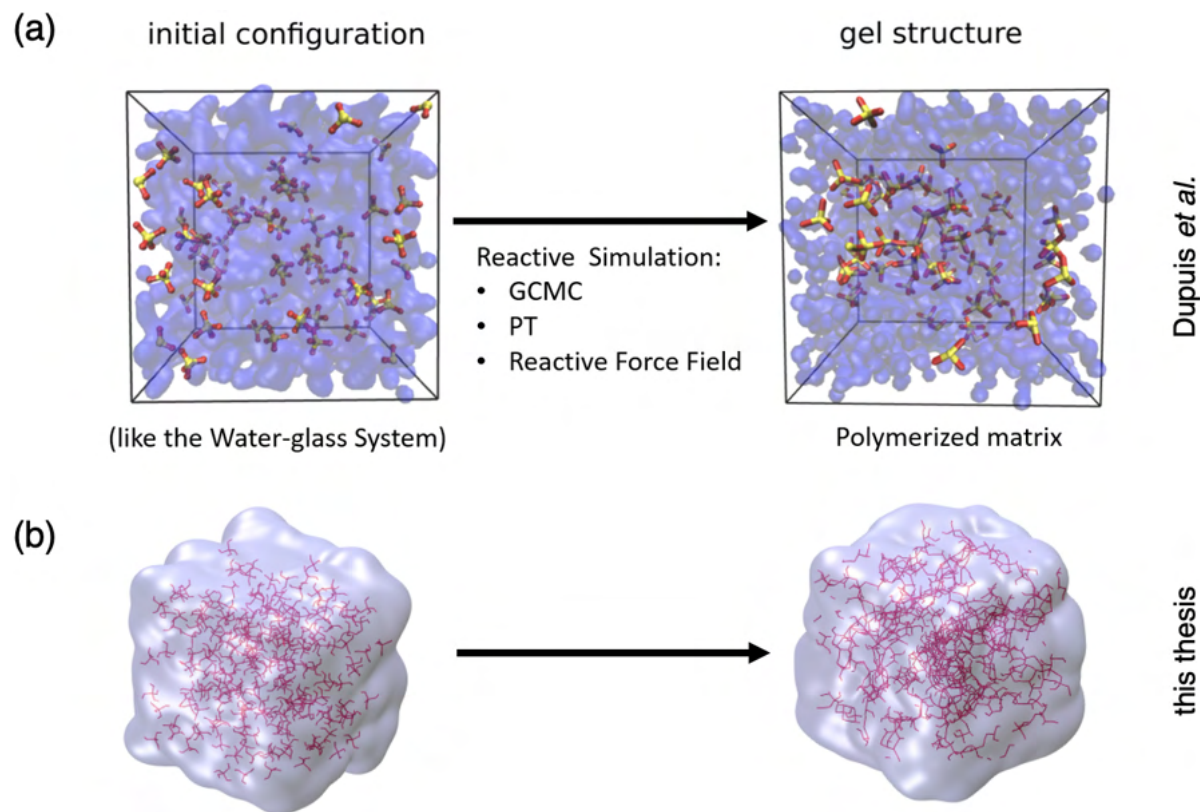


Figure 3.20: **(a)** Scheme utilized by Dupuis *et al.* [173] to reproduce the matrix formation process starting from an initial condition identical to the Waterglass system. Water molecules are represented by the blue cloud, while orthosilicic acid molecules and chains by the typical red-and-yellow ball-and-stick representations. Small chains showing some branches can be distinguished in the polymerized matrix. **(b)** Reproduction of the Dupuis *et al.* simulation protocol. In the image there are shown the first and final frames of the same simulation performed by me: red molecules in the left box are of Si(OH)_4 units, while in the right a connected, polymerized matrix was generated during a single Reactive MD trajectory with ReaxFF. Details of my simulations can be found in the Appendix A. Adapted from Ref. [173].

This particular scheme was utilized for the following reasons: the siloxane bond formation between Si(OH)_4 molecules is intrinsically slow due to high energy barriers, making the spontaneous formation of a silica network computationally expensive at normal temperatures. PT allows replicas at different temperatures to exchange configurations, thereby overcoming such kinetic barriers without artificially heating the system to unrealistic temperatures. In contrast, other studies using the Fogarty *et al.* reactive force field without PT sometimes relied

on extremely high temperatures to induce faster fusion of the silica network [180], which may distort the dynamics compared to realistic gelation conditions. Therefore, PT provides a more controlled and physically meaningful route to sample siloxane bond formation and gelation in silicate systems.

Proteins and Silica Polymerization within the ReaxFF Framework

However, the main goal of this section should be to reproduce the effect that the forming matrix might have on a protein structure, not really reproducing polymerization itself. ReaxFF has proven highly effective in modeling reactive processes in silica systems, but the simultaneous description of protein–silica–water systems remains an open challenge. A major limitation is the lack of a unified parameter set capable of accurately capturing:

- silica network formation,
- protein structure and amino acid chemistry,
- their mutual interactions in aqueous environments.

While separate parameterizations exist for silica–water and protein–water systems, their combination is not feasible nowadays, as they belong to very different sub-branches. The direct merging of different ReaxFF parameter sets is not considered reliable, as it breaks the internal consistency of the force field (merging different parameters is not a good idea). Current approaches typically model silica and biomolecular subsystems independently. The full coupling between these components –necessary to describe biosilicification or APT embedding processes– cannot yet be robustly captured. Although some efforts have been made toward unified parameterizations, no broadly accepted ReaxFF force field for protein–silica systems is currently available. Therefore, another more successful strategy, described in the next section, could be used to overcome both the ReaxFF parametrization issue and also the fact that silica polymerization should occur in greater timescales.

3.6.2 Coarse Grained Molecular Dynamics and Reactivity: A New Hope

In this section, I will provide a brief overview of the results obtained mainly by my colleague Francesco Carnovale within the MIMOSA project (I contributed only marginally). His work focused on studying the late stages of the encapsulation process through Coarse-Grained (CG) Molecular Dynamics Simulations. In Chapter 2.3 there is a brief and naive explanation of the CG MD framework.

As described in the previous sections, atomistic MD (classical, reactive) cannot provide a proper pipeline to investigate the late stages of the polymerization process, i.e. a forming SiO_2

matrix around the protein structure. In fact, the simulation of silica polymerization represents a formidable challenge in computational chemistry, requiring a balance between the chemical accuracy of bond formation and the long-timescale dynamics of gelation. This section connects coarse-grained with the current discussion, beginning with the description of an initial CG model of SiO_2 polymerization, named Sticky-MARTINI, published in 2022 (Ref. [181]), that was further refined by Francesco Carnovale, resulting into a final model, called Si8B MARTINI 3. More detailed information about his work can be found in his PhD thesis [182].

The Sticky-MARTINI Model Within the MARTINI 2 Framework

The Sticky-MARTINI model, introduced by Carvalho *et al.* (2022) [181], was designed to simulate the polymerization of silicic acid ($\text{Si}(\text{OH})_4$) into complex Q^n species, where n defines the degree of polymerization, meaning that 0, 1, 2 identify three separate species, $\text{Si}(\text{OH})_4$, $\text{Si}_2\text{O}_7\text{H}_6$, $\text{Si}_3\text{O}_{10}\text{H}_8$, respectively. Based on the Martini 2 force field, this model utilized a multi-bead sticky particle approach. Each silica unit consisted of a central virtual site (representing the Silicon atom, Si) surrounded by four Acceptor-Donor (AD) beads representing the hydroxyl groups and four Repulsive-Virtual (RV) beads intended to maintain tetrahedral geometry, as shown in Fig. 3.21. These two particle types might allow to reproduce the correct behavior of the bond between two or more $\text{Si}(\text{OH})_4$ units [181].

From Chemical Reactivity to Effective Interactions: Defining Reactivity in the Context of CG MD Simulations

Some words should be spent here to define some differences between what Sticky-MARTINI truly models and chemical reactivity. In fact, a fundamental limitation of standard coarse-grained (CG) force fields, including MARTINI, is that they are inherently non-reactive: the molecular topology is fixed, and chemical bonds cannot be formed or broken during the simulation. In fact, performing classical CG simulations involving a protein (Ubiquitin) and silica species ($\text{Si}(\text{OH})_4$, $\text{Si}_2\text{O}_7\text{H}_6$ and $\text{Si}_3\text{O}_{10}\text{H}_8$) is possible, as parameters for the CG beads that represent the two entities are already present in the standard (thus, **non-sticky**) MARTINI force field. Fig. 3.22 reports results from classical CG simulations, that can be seen as CG versions of the classical atomistic simulations presented in the previous chapters (specifically the Mixed scenario). Three standard, different types of beads were utilized to model silica, namely N5, P4, P6. These results qualitatively resemble somehow what found in Figs. 3.9 and 3.11. However, the resolution of the system is too low to infer the same conclusions, and moreover these simulations, still displaying a classical nature, cannot capture the silica polymerization.

This poses a challenge for processes such as polymerization, where the formation of new covalent bonds is the central physical mechanism. To address this limitation within the classical MAR-

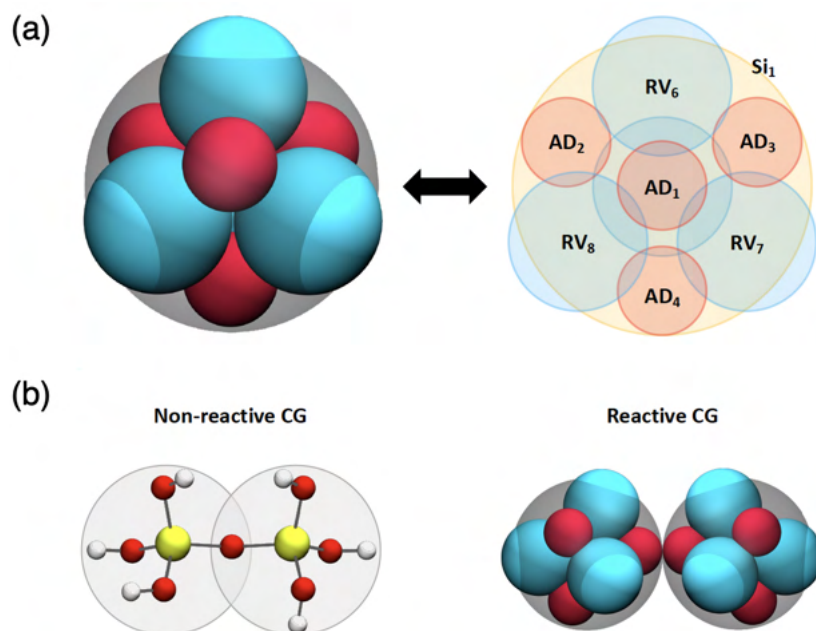


Figure 3.21: (a) The sticky-MARTINI bead organization: each silica unit is represented by a central virtual site (Si), four Acceptor–Donor (AD) beads modeling the hydroxyl groups, and four Repulsive–Virtual (RV) beads that enforce the tetrahedral geometry. The AD beads act as reactive sites via attractive interactions, while the RV beads maintain the correct spatial arrangement. (b) Comparison between a $Si_2O_7H_6$ molecule in the non-reactive CG framework (the two intersecting beads are represented by gray spheres) and in a reactive CG framework. Adapted from Ref. [182].

TINI forcefield, extensions such as the *Sticky MARTINI* model have been developed, enabling an effective representation of reactivity through bead association rather than explicit bond formation. In the context of silica polymerization, this approach provides a practical route to capture network formation while retaining the computational efficiency of coarse-grained simulations.

At the atomistic level, silica polymerization involves condensation reactions between silanol groups, leading to the formation of Si–O–Si linkages and the release of small molecules (e.g., water). These processes are governed by complex reaction pathways, including bond breaking, bond formation, and solvent-mediated effects.

In a coarse-grained representation, such detailed chemistry is not explicitly resolved. Instead, the key idea is to map reactive chemical groups onto CG beads and to encode their propensity to form bonds into effective, short-range attractive interactions. In *Sticky MARTINI*, this is achieved by introducing specific bead types – the multi-bead sticky particle approach aforementioned – that interact *via* enhanced Lennard–Jones potentials. These interactions are tuned such that, when two reactive beads come into proximity, they experience a sufficiently

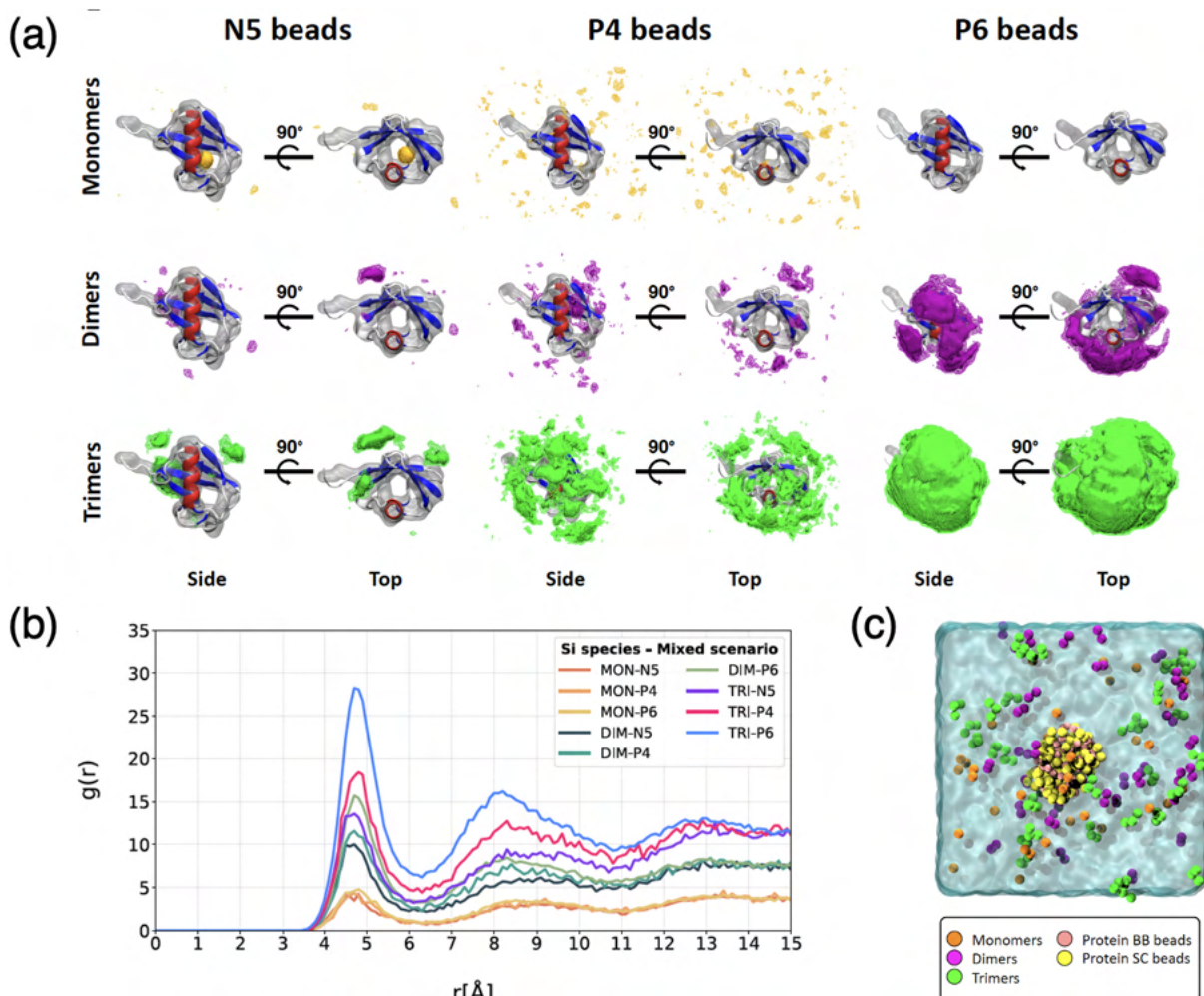


Figure 3.22: Classical Coarse-Grained MD Simulations Replicating the Mixed System. (a) 3D density maps of silica species around the surface of the ubiquitin protein in classical CG MARTINI simulations. Results resemble that ones of Fig. 3.11, and are shown for three different classical $\text{Si}(\text{OH})_4$ bead types, namely N5, P4, P6. (b) Radial distribution functions $\rho(r)$ calculated for monomers, dimers, and trimers using three different silica bead types (N5, P4, and P6). These values are computed following the same procedure adopted for the all-atom simulations shown in Fig. 3.9. (c) Rendering of the mixed system: ubiquitin (yellow and pink beads) in aqueous solution (water represented as cyan surface) at 150 mM NaCl concentration (ions not shown), with orthosilicic monomers, dimers, and trimers depicted in orange, violet, and green, respectively. Adapted from Ref. [182].

strong attraction to form a stable association. Importantly, this association is not a true covalent bond: it does not involve a change in topology, and it remains reversible on sufficiently long time scales. Therefore, the central approximation of the Sticky MARTINI approach is to treat bead

association as a proxy for chemical bond formation. This mapping relies on the assumption that, for the purposes of mesoscale structure formation, the presence of a persistent contact between two reactive sites is functionally equivalent to a covalent linkage. To implement this, the interaction strength between sticky beads is significantly increased compared to standard MARTINI interactions, and the interaction range is often adjusted to favor contact formation. As a result, once two reactive beads associate, they remain in close proximity for extended periods, effectively mimicking the permanence of a chemical bond.

The use of sticky interactions introduces an implicit description of both thermodynamics and kinetics. The depth of the interaction potential controls the equilibrium constant of association, thereby encoding the thermodynamic favorability of bond formation. At the same time, the diffusive encounter rate between beads governs the effective reaction kinetics.

However, it is important to emphasize that this mapping is approximate. The resulting association rates do not correspond directly to chemical reaction rates, and the reversibility of bead associations may not reflect the irreversibility of certain polymerization reactions. As such, parameterization often involves tuning interaction strengths to reproduce macroscopic observables, such as degree of polymerization or network structure, rather than microscopic reaction mechanisms.

This approach allows one to study the evolution of mesoscale structure over time scales and system sizes that would be inaccessible to reactive atomistic simulations. At the same time, it preserves the simplicity and transferability of the MARTINI framework. Despite these limitations, the method provides a valuable compromise between chemical realism and computational efficiency, enabling the study of reactive processes within a coarse-grained paradigm.

Limitations of Sticky-MARTINI

While Sticky-MARTINI was a breakthrough in observing microsecond-scale gelation, it suffered from three fundamental physical discrepancies that Ref. [182] seeks to rectify:

1. **Geometric Distortion:** The equilibrium Si–Si distance in the Carvalho model was approximately 4.82 Å. This is significantly higher than the experimental values for amorphous silica, which typically range between 3.1 and 3.6 Å. This "stretched" network fails to accurately capture the density and porosity of real silica gels.
2. **Coordination Artifacts:** The model frequently allowed "triple bonds" or over-coordination, where multiple AD beads clustered at a single site due to insufficient steric shielding by the RV beads.
3. **Numerical Instability:** To prevent unphysical bead overlaps and simulation crashes, the model required an integration time step of 6 fs or less, which is computationally expensive

for coarse-grained (CG) regimes [181].

Moreover, the whole model was developed under the MARTINI 2 force field, that is known to have problems in modeling proteins. For this reason, the work performed by Francesco was mainly oriented in transfer the Sticky-MARTINI implementation to MARTINI 3 [183]. Such detail has an extremely high relevance, as this means that this model developed by my colleague is compatible with the parametrization of proteins within MARTINI 3. In other words, this solves the issue that ReaxFF had before, enabling the representation of a polymerizing matrix around a protein structure within a single simulation run.

From Sticky-MARTINI to Si8B MARTINI 3 Model

The reactive framework of Sticky MARTINI should be brought into the **Martini 3** ecosystem [183], which offers superior resolution through Tiny (T) and Small (S) beads instead of regular (R) size beads. Three subsequent models were introduced by Francesco Carnovale, each designed to overcome the shortcomings of the previous one. Here I summarize the first two, while the final refined third model will be described in the following section (see also Fig. 3.23):

- **Model 1** was a direct port of the Carvalho geometry into Martini 3. While it maintained stability, it inherited the unphysical 4.82 Å Si-Si distance present i Sticky-MARTINI. More critically, Model 1 exhibited a "freezing" phenomenon, where the system would undergo an unphysical phase transition into a quasi-crystalline lattice rather than remaining an amorphous gel.
- **Model 2** attempted to correct the Si-Si distance to 3.82 Å by protruding the AD beads further from the central Si site. However, this protrusion exposed the reactive sites, leading to a massive increase in over-coordination and "triple-bound" artifacts, rendering the resulting network chemically inaccurate.

Model 3: the Si8B MARTINI 3 Parametrization

First of all, the name Si8B is a descriptive shorthand for the specific geometry and composition of the coarse-grained silica unit utilized. In fact, the naming convention breaks down as follows: Si (Silicon) represents the central silicon atom of the orthosilicic acid monomer $\text{Si}(\text{OH})_4$. In this model, the silicon is specifically a massless virtual site (using the Martini 3 TN6 bead type) positioned at the geometric center of the unit to ensure structural stability. The 8 (Eight Peripheral Sites) refers to the eight distinct interaction sites that surround the central silicon to manage its chemical stickiness and geometry: 4 Attractive Dummy (AD) beads –representing the four hydroxyl ($-\text{OH}$) groups and are responsible for forming the siloxane bonds ($\text{Si}-\text{O}-\text{Si}$)

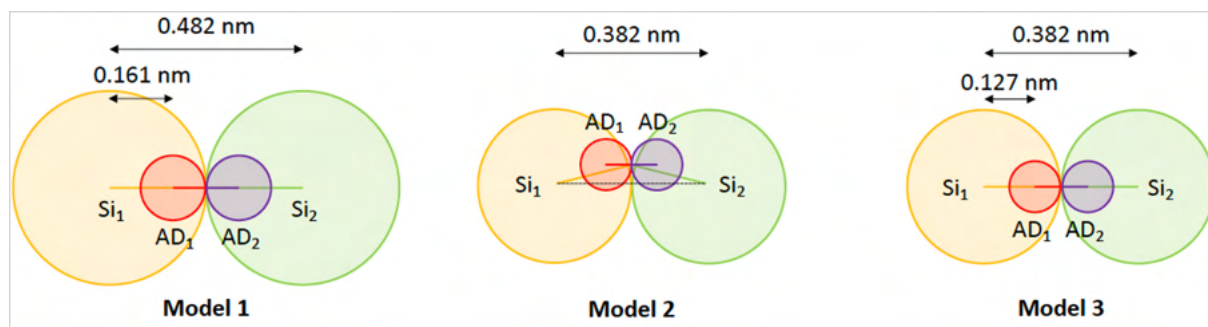


Figure 3.23: The geometries of the three Si8B bead models developed in this work are illustrated. Model 1 is shown slightly larger to emphasize that it uses standard-sized beads rather than tiny Martini beads. The beads are in scale within the drawing of the same model but not between the different models. Each model shows the interaction between two Si8B beads; for clarity, only the central Si atom and a single AD bead of each interacting Si8B bead are represented.. Adapted from Ref. [182].

through strong attractive potentials– and 4 Repulsive Virtual (RV) sites –positioned diametrically opposite to the AD beads to create a star-shaped octahedral configuration, as in Fig. 3.21, providing steric repulsion, preventing the unit from forming more than four bonds (avoiding unphysical over-coordination). Lastly, the letter B (Bead) refers to the fact that this entire 9-point construction (1 central Si + 8 peripheral sites) is treated as a single, unified Si8B bead in the coarse-grained simulation framework. Model 3 is presented as the definitive solution to the flaws of its predecessors and the original Sticky-MARTINI model.

Geometric Precision and Si–Si Bond Length The primary achievement of Model 3 is the reduction of the Si–Si distance to a physically relevant 3.82 Å. This was achieved by utilizing the smaller footprint of Martini 3 TN6 beads, allowing the reactive sites to sit closer to the silicon center without overlapping. This 20% reduction in bond length compared to the Carvalho 2022 model [181] allows for a much more realistic description of the silica during the polymerization of the matrix.

The Physics of Repulsion (ϵ_{RV}) and Coordination Control The parameter responsible for the repulsion energy ϵ_{RV} was also modified within Model3. In Sticky-MARTINI, the repulsion was often insufficient to prevent over-coordination, while Model 3 utilizes a robust repulsion value of $\epsilon_{RV} = 44$ kJ/mol. This value was meticulously tuned to provide a steric exclusion zone around each reactive AD site. At lower ϵ_{RV} values, the model reverts to the failures of Model 2, allowing unphysical situations like the formation of triple bonds. At the optimized 44 kJ/mol, the model strictly enforces tetrahedral connectivity, ensuring that each AD site can only interact with a

single partner, thereby mimicking the valency of real oxygen bridges in $Si - O - Si$ bonds.

Angular Distribution and Amorphous Stability Unlike Model 1, which suffered from unphysical crystallization, Model 3 produces a broad Si–O–Si angular distribution peaking at approximately 165° . While slightly higher than the experimental 145° for bulk silica, this angle provides the necessary topological flexibility for the CG beads to form a disordered, glass-like network. This is essential for simulating the random growth of silica clusters.

Kinetics and Temporal Stability In Model 3, the interaction energy between reactive sites is set to $\epsilon_{AD} = 50$ kJ/mol. This high energy ensures that once a bond is formed, it is effectively irreversible on the simulation timescale, accurately reflecting the covalent nature of siloxane bonds. Despite the high forces involved, the model remains numerically stable at a 5 fs time step, avoiding the simulation crashes observed in the Carvalho model [181].

The Relevance of Backmapping to Atomistic Resolution

Before introducing the results of Model 3, it is first essential to report a fundamental concept within CG simulations, the backmapping procedures. While the sticky-bead approach enables the simulation of silica polymerization at mesoscopic scales, the resulting structures lack explicit chemical detail. In particular, the representation of reactivity through bead association does not uniquely define the underlying atomistic connectivity, bond geometries, or local chemical environments. For this reason, backmapping to an atomistic description becomes a crucial step for recovering chemically resolved structures.

Backmapping consists in reconstructing atomistic coordinates from a coarse-grained configuration by assigning atomistic fragments to each bead and restoring the molecular topology. In the context of CG simulations and MARTINI, this step is non-trivial: because bead association does not correspond to explicit bond formation, additional criteria must be introduced to infer which contacts should be converted into covalent linkages. Typically, distance-based thresholds and geometric considerations are used to identify persistent associations and translate them into bonded atomistic structures.

The relevance of backmapping is twofold. First, it enables structural refinement: once an atomistic model is reconstructed, energy minimization and short atomistic MD simulations can relax local geometries, yielding physically realistic bond lengths, angles, and coordination environments. This is particularly important for silica systems, where network connectivity and local tetrahedral order are key structural features. Second, backmapping provides access to properties that cannot be described at the coarse-grained level, such as detailed bonding patterns, defect structures, vibrational properties, and spectroscopic observables. It also allows

one to validate the coarse-grained model a posteriori, by comparing reconstructed atomistic structures with experimental data or fully atomistic simulations.

Overall, backmapping serves as a bridge between the efficiency of coarse-grained simulations and the chemical accuracy of atomistic models. In the case of effective reactive CG approaches such as Sticky MARTINI, it is an essential step for interpreting the resulting structures in chemically meaningful terms.

The Late Stages of Silica Polymerization

Model 3 reproduces the SiO_2 polymerization Fig. 3.24 reports the results from Model 3. First of all, the number of Q^3 and Q^4 species (i.e. silica species that combine 4 or 5 $\text{Si}(\text{OH})_4$ monomers) increases over the simulation, as can be seen by points (a) and (b). This indicates that the silica is properly polymerizing, forming a interconnected matrix structure. In fact, over longer timescales the system evolves into an amorphous SiO_2 network (see point (c)). This transition does not occur immediately, but only after a few hundreds of nanoseconds (approximately 200 ns in most replicas). At that stage, a solid matrix surrounding the protein has formed, enabling backmapping and subsequent atomistic investigation of the system. In conclusion, by coupling the advanced bead types of Martini 3 with a tangent geometric arrangement, Model 3 represents a significant advancement over the 2022 Sticky-MARTINI paper. It corrects the fundamental Si–Si distance error, eliminates unphysical over-coordination through tuned ϵ_{RV} parameters, and provides the stability required for large-scale MD studies of protein-silica interactions.

Modeling the Silica Polymerization around Protein Structures As mentioned previously, the main goal of this work should be to utilize the faster dynamics of CG coupled with an effective reactive framework to study the formation of the matrix around the protein structure. This is a non-trivial, but achievable task, since the parametrization of Model 3 lies within the MARTINI 3 force field, thus compatible with aminoacidic beads. However, up to now Model 3 models only silica and water systems; no protein was inserted within the simulations described in Fig. 3.24. In fact, to do so the Model 3 should be partially re-tuned, in order to account for the interactions with other MARTINI 3 beads that are typically used to represent biomolecules within CG frameworks. This is a process that is affordable and achievable, but it requires time and effort. Francesco, me and the whole MIMOSA researchers are confident that such implementation will be pursued successfully in the next upcoming months.

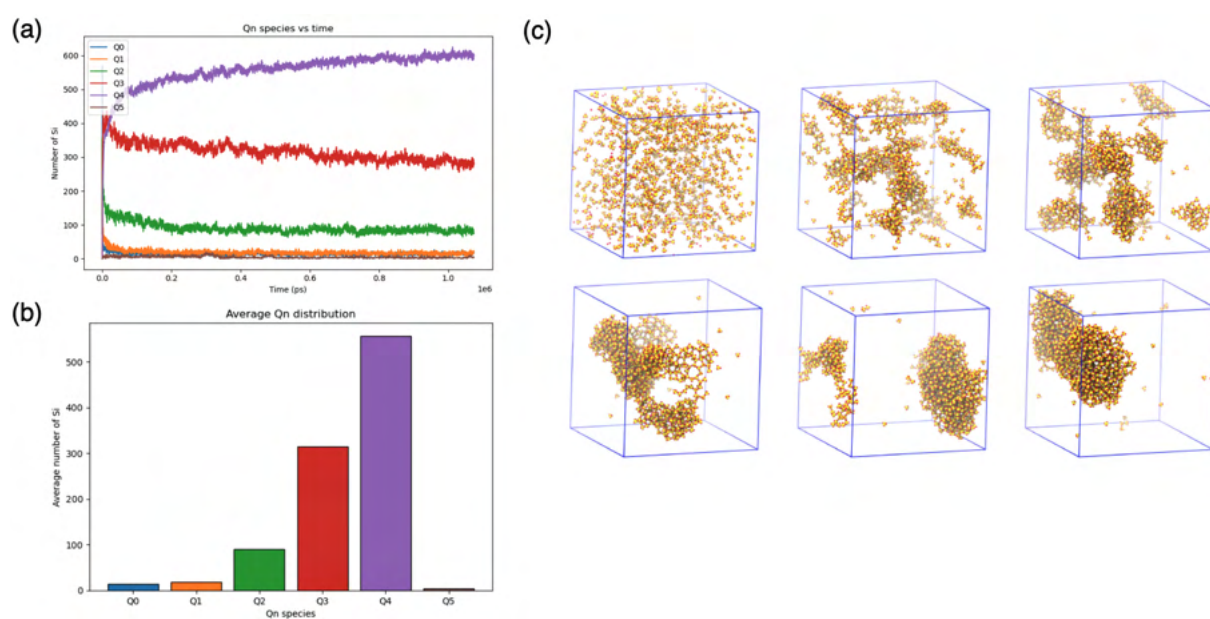


Figure 3.24: Results of Reactive CG Simulations of Model 3 (a) Number of polymerizing silica species Q^n in function of time, where n is the degree of polymerization starting from 0 –i. e. Q^0 corresponds $\text{Si}(\text{OH})_4$. (b) Average number of polymerizing silica species Q^n across the whole simulation. (c) Schematic representation of one simulation trajectory (six frames are shown) representing a polymerizing silica matrix; first frame is the upper left, last frame is lower right (please read first row first, then second row). Adapted from Ref. [182].

Unraveling THz-mediated Ion Field Evaporation in APT via *ab initio* Molecular Dynamics

This Chapter is primarily inspired by Ref. [184]. The opening section contains a brief introduction to Terahertz-assisted Atom Probe Tomography. A second section describes recent experimental advances in the application of such laser type on the APT analysis of insulating samples like SiO₂ matrices, in correlation to a relevant aspect, called THz pulse polarity. These recent outcomes were directly produced by the group of Angela Vella from the University of Rouen (France), collaborators within the MIMOSA project. Such section establishes the foundational framework for all the following ones, motivating the use of *ab initio* MD simulations to investigate essential microscopical details, that provide a theoretical baseline needed for the understanding of these found experimental results. A coupled scheme –Real time, time-dependent density functional theory (RT-TDDFT) and Ehrenfest dynamics (ED)– was employed to reproduce *in silico* the THz-APT experimental setup dynamics. To recall, more information about RT-TDDFT and ED can be found in Chapter 2, Sections 2.5.2 and 2.5.3, respectively.

4.1 Introduction to THz-assisted Atom Probe Tomography

4.1.1 THz Radiation Properties

Terahertz (THz) radiation occupies the spectral region between microwaves and infrared light, spanning approximately $0.1 \lesssim \nu \lesssim 30$ THz, corresponding to wavelengths from about 3 mm down to $30 \mu\text{m}$. A useful reference point is $\nu = 1$ THz, which gives a period $\tau = 1$ ps, wavelength $\lambda \approx 300 \mu\text{m}$, and photon energy $E_\gamma \approx 4.1$ meV (see Fig. 4.1)

Because these photon energies are far below typical electronic band gaps, THz radiation is generally *non-ionizing* and induces significantly less heating than UV or visible light. Many dry dielectric materials are relatively transparent in this frequency range, while water exhibits strong absorption [186, 187]. In the time domain, THz pulses are typically *single-cycle* transients with durations in the sub- to few-picosecond range, comparable to ion evaporation times in atom probe tomography (APT). Here, the single-cycle term refers to an extremely short burst of

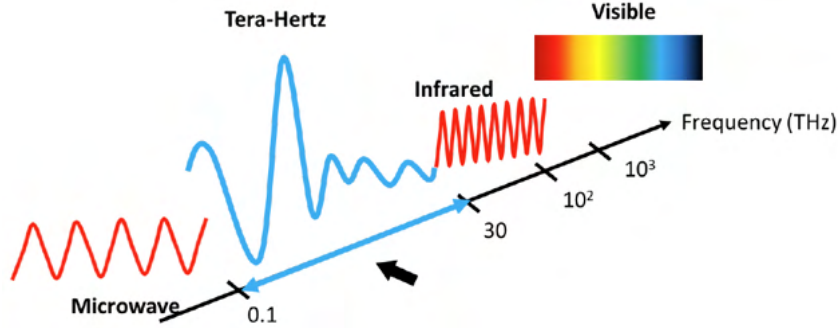


Figure 4.1: The Terahertz Laser Frequency. Electromagnetic spectrum highlighting the THz band (shaded): $0.1 \lesssim \nu \lesssim 30$ THz, $\lambda \approx 3$ mm down to 30μ . At 1 THz, the period is 1 ps and the photon energy is ≈ 4.1 meV. Adapted from Ref. [185]

radiation, whose the relative associated field oscillates only once (one cycle). Such pulses are very fast, precise and strong at their peak.

Free-space THz sources can generate peak fields on the order of tens to hundreds of kV/cm. However, in the context of APT there is a subtle detail: in fact, at nanosharp emitters, these fields can be *significantly enhanced* through near-field effects arising from geometric field concentration and antenna-like coupling at the tip apex [188, 189, 190]. In practice, the effective field at the apex is determined *not only* by the incident THz amplitude, but also by the *emitter geometry* (apex radius and shank angle), electrostatic boundary conditions, and applied DC bias. These factors collectively define the local field enhancement, that plays a pivotal role within APT frameworks. A more detailed discussion of these mechanisms follows in the next sections. This is a crucial step for the comprehension of how THz APT can evaporate ions.

4.1.2 THz-assisted Field Evaporation and Thermal Effects

THz excitation modifies the emission physics in APT. The research group of Angela Vella, a collaborator within the MIMOSA Consortium, performed an APT experiment on aluminum tips, trying to compare a single-cycle THz pulse and a near-infrared (NIR) pulse in terms of variable delay and evaporation rate [93]. Specifically, the two pulses were sequentially in two ways: THz first and NIR last, and NIR first and Thz last. Notably, the setup is built so that evaporation is triggered only if the two pulses energies sum up.

Two different behaviors were observed. (i) If THz precedes NIR, no evaporation is observed, i.e. THz does not store energy in the system and acts only during its instantaneous presence and when NIR arrives, only its energy is not enough to cause evaporation. (ii) If NIR precedes THz, as THz arrives a strong increase in the evaporation rate is observed (the two pulses sum

up together). Thus, NIR must deposit energy into the system, leaving a persistent effect over picoseconds, and the only relevant quantity that persists on that timescale is temperature. Thus, this asymmetry indicates that NIR excitation primarily induces thermal effects (heats the apex, thermal assistance), whereas THz excitation promotes mostly athermal, field-driven emission (only during its presence) [93].

As seen by equation 1.11, the evaporation rate depends on both the lattice temperature T and the activation barrier $Q(F)$, which decreases with increasing electric field F (see equations 1.13 and 1.14). NIR pulses mainly increase T through electronic and lattice heating, while THz pulses predominantly increase F with negligible heating [191]. Thus, this simple model treating NIR as a ΔT perturbation and THz as a ΔF perturbation successfully supports a field-driven evaporation mechanism for THz [93]. As a result, this field-dominated evaporation regime reduces thermal artifacts and improves resolution, suppressing thermal tailing in the relative APT spectra.

4.1.3 THz-assisted APT and Field Enhancement

Field evaporation is a ionization-driven mechanism that requires apex fields of several V/nm. To reach these values, THz-assisted APT exploits strong near-field *antenna* enhancement: when $\lambda_{\text{THz}} \gg R_{\text{tip}}$, the local field can be amplified by factors up to 10^3 . The total field can then be written as

$$F_{\text{tot}}(t) = F_{\text{DC}} + \eta F_{\text{THz}}(t). \quad (4.1)$$

where F_{DC} is the static electric field due to the applied bias, and $F_{\text{THz}}(t)$ is the time-dependent THz field in free space (not enhanced). The factor η represents the near-field enhancement at the tip apex, arising from geometric field concentration and antenna-like effects. Physically, this expression indicates that the local field experienced by surface atoms is the superposition of a constant bias field and a transient contribution from the THz pulse, which is strongly amplified at the nanoscale (needle apex). Although the incident THz field is relatively weak, the enhancement factor $\eta \gg 1$ can increase its modulus at the apex by several orders of magnitude. The total field can transiently exceed the evaporation threshold during the THz pulse, even when the DC field alone is subcritical. This enables ultrafast, field-driven evaporation within a picosecond time window, after which the field returns to its baseline value F_{DC} . This enhancement is thought to be *material-dependent*: metallic tips typically exhibit large enhancement factors due to conduction electron dynamics, whereas semiconductors show a modest amplification.

While equation 4.1 describes the origin of the total apex field *via* near-field enhancement, there is a practical experimental relation used to extract the THz field enhanced contribution $\eta F_{\text{THz}}(t)$ from a measured voltage, also reported in equation 1.18. Applied to the THz case,

$\eta F_{\text{THz}}(t)$ is then given by:

$$\eta F_{\text{THz}}(t) = \frac{V_{\text{THz}}(t)}{k_f R_{\text{tip}}}, \quad (4.2)$$

where R_{tip} is the apex radius and k_f is a geometry factor. Using this approach, enhancement factors as high as $\eta \approx 2 \times 10^3$ have been reported.

The magnitude of the THz field at the tip apex, η , is governed by the ratio between the radiation wavelength λ and the tip radius R_{tip} . For THz radiation, $\lambda_{\text{THz}} \sim 10^2\text{--}10^3 \mu\text{m}$, while $R_{\text{tip}} \sim 10\text{--}100 \text{ nm}$, such that $\lambda_{\text{THz}} \gg R_{\text{tip}}$. In this regime, the tip behaves as a subwavelength metallic antenna: the incident electromagnetic wave drives collective motion of conduction electrons, leading to charge accumulation at the apex and strong localization of the electric field (up to $\eta \sim 10^3$, as aforementioned).

In contrast, for UV radiation $\lambda_{\text{UV}} \sim 10^2 \text{ nm}$, which is comparable to the tip radius: $\lambda_{\text{UV}} \sim R_{\text{tip}}$. Under these conditions, the quasi-static approximation breaks down and the electron motion cannot follow the oscillating field coherently across the structure. Consequently, no significant charge build-up occurs at the apex, and the antenna effect is strongly suppressed in the case of UV, leading to $\eta \approx 1$.

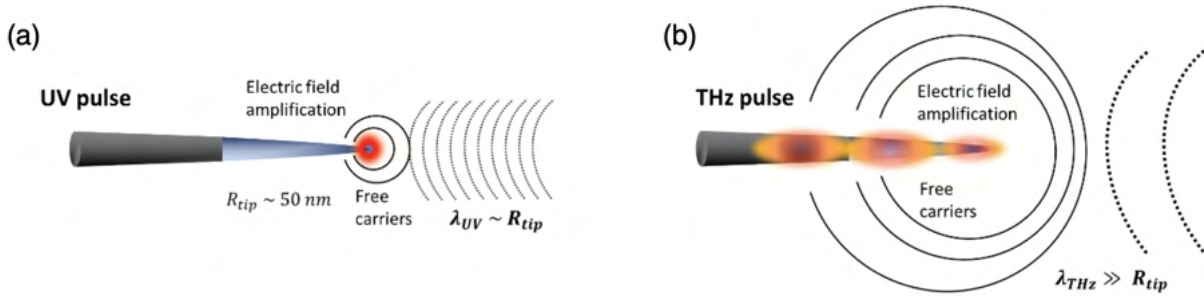


Figure 4.2: Near-field amplification at a nanotip. Comparison between UV ($\lambda_{\text{UV}} \sim R_{\text{tip}}$) (a) and THz ($\lambda_{\text{THz}} \gg R_{\text{tip}}$) (b) illumination. In the UV regime, the wavelength is comparable to the tip size, so the field interacts with the detailed geometry of the apex and produces only weak enhancement; in the THz regime the wavelength is much larger than the tip, causing it to behave like a small antenna that concentrates charge at the apex and generates a strongly enhanced, localized electric field. Adapted from Vella *et al.* [93].

To sum up, THz radiation remains non-ionizing in terms of photon energy, but the antenna enhancement through the tip apex generates extremely strong local electric fields that induce field ionization and evaporation *via* barrier lowering mechanisms. Moreover, while THz radiation can be efficiently concentrated at the tip apex through this field enhancement, UV excitation primarily interacts with the material through absorption and electronic excitation, without substantial field amplification. In other terms, UV-assisted APT ionizes by kicking electrons

out of the sample (high-energy photons), THz-assisted APT by pulling them out of the tip with a strong total field.

4.1.4 The Single-Cycle THz Pulse Polarity in APT

A defining feature of a single-cycle THz pulse is its temporal field profile, commonly referred to as its *polarity*, which describes the instantaneous direction of the field within the pulse. A THz pulse may exhibit positive polarity, where the field points predominantly in one direction, negative polarity, where the field points oppositely, or bipolar behavior, where the field reverses direction during the pulse. In the context of atom probe tomography (APT), the polarity of the THz pulse is a crucial parameter because the local electric field at the apex of the sample governs the emission of ions and electrons. A unipolar positive pulse can preferentially enhance the extraction of negatively charged electrons, while a negative pulse may favor the emission of positive ions. Bipolar pulses, by contrast, may partially cancel the field-driven emission in one direction or enable sequential emission of opposite charges, depending on the precise temporal overlap of the field extrema with the atomic evaporation process. Fig. 4.3 reports examples of THz positive (a) and negative (b) polarity pulses.

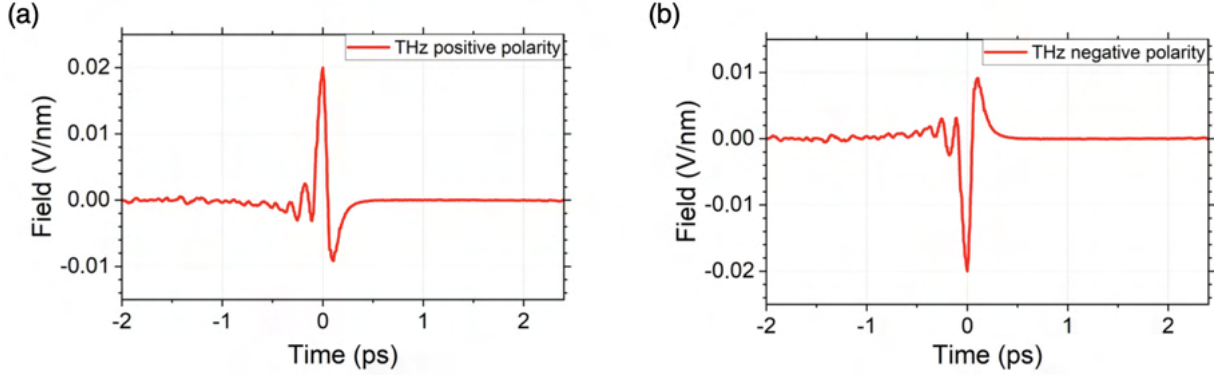


Figure 4.3: Electro-optical traces of a positive polarity (a) and a negative polarity (b) THz pulse. Adapted from Ref. [184].

This polarity-dependent effect is particularly important in THz-assisted APT, where field-driven ionization dominates over thermal mechanisms due to the ultrashort pulse duration. The temporal asymmetry introduced by the polarity can modify the local potential barrier at the apex on sub-picosecond timescales, effectively controlling which species are emitted and when. Moreover, the shape of the pulse—including the relative amplitudes, duration, and polarity sequence—can be used to tune the field-enhancement dynamics, making THz pulses versatile tools for selective ion emission. Despite these fundamental implications, the influence of THz polarity on the emis-

sion from nonconducting or dielectric nanostructures, and on the overall ionization dynamics in APT, remains largely unexplored. Understanding and controlling polarity effects therefore represents a key avenue for optimizing THz-driven APT and for extending its capability to a broader range of materials.

4.2 Motivating *ab initio* MD Simulations: Anomalous THz Pulse Polarity Behavior in THz-assisted APT of Silica Tips

Recently, the group of Angela Vella from the University of Rouen (France) made a huge leap forward towards the comprehension of the THz laser in the context of APT, extending its usage also to amorphous silica (SiO_2) tips [184]. Silica is an insulator material with a large band gap of about 9 eV and low electrical conductivity. As mentioned above, THz lasers work really well for metals, as the tip geometry can easily enhance the THz field, resulting into a good evaporation trend. However, such enhancement depends rigorously on the utilized materials, and is known to decrease when the probed specimen is a semiconductor or an insulator. Moreover, also the THz laser polarity might play a pivotal role in this context. In fact, THz polarity can significantly influence the interaction between THz light and material during APT experiments, especially with regard to the emission of positive or negative charges.

Thus, this work by Vella *et al.* [184] presents for the first time the single-cycle evaporation of cations from nonconducting amorphous SiO_2 matrices using THz pulses, emphasizing the importance of pulse polarity for the emission process. The results were also compared with outcomes coming from LaB₆ semi-metallic samples, here taken as a benchmark. Specifically, two separate analyses were carried out, one on LaB₆ and the other on silica specimens. For each case (SiO_2 , LaB₆), two separate THz single-cycle pulses with an asymmetrically large positive or negative peak field were utilized, as in Fig. 4.3; therefore, four final APT analyses were conducted, in order to monitor and compare the pulse polarity behavior across semi-metallic and insulating systems.

LaB₆ tips were fabricated from single-crystal rods via electrochemical etching followed by focused ion beam (FIB) milling. Silica samples were prepared according to a protocol originally developed for protein encapsulation in silica for APT analysis [82, 115], as described in Chapter 1, with the only modification being the absence of proteins. Additional details on sample preparation are provided in Ref. [192]. A representative scanning electron microscopy image is shown in the Appendix B, Fig. B.1 (a), highlighting a $\sim 5 \mu\text{m}$ long silica tip mounted and glued on a tungsten pre-tip. Spatially resolved mass spectra were acquired using a THz-assisted atom probe tomography (APT) setup, schematically shown in Fig. B.1 (b).

The needle-shaped specimen is biased with a DC voltage between -0.8 and 20 kV, gen-

erating an enhanced apex electric field exceeding 10 V/nm. Field evaporation is triggered by the combined action of this static field and single-cycle THz pulses, which are generated *via* a two-color laser-induced plasma [193, 194]. This means that the static field alone *cannot* induce evaporation. Although the incident THz field is on the order of 0.02 V/nm, near-field enhancement at the tip apex amplifies it by roughly three orders of magnitude [195], resulting in peak fields between 5 and 20 V/nm. Overall, the static and THz fields lie in comparable ranges of approximately 10–20 V/nm and 5–20 V/nm, respectively.

The THz laser system consists of the fundamental (800 nm) and its second harmonic, producing asymmetric electron currents that emit THz radiation. The final pulse energy (non-enhanced) on the order of $\sim 0.1 \mu\text{J}$ ($\approx 10^{-7}$ J), and the relative laser frequencies are in the range of 1–10 THz. These raw pulse energy will be enhanced by the APT instrument. The pulse repetition rate is 1 kHz, and the THz field amplitude is controlled through the input power of the fundamental harmonic. The emitted THz pulses are characterized by electro-optical sampling in a nonlinear crystal, with representative traces of positive and negative pulses shown in Fig. 4.3. The beam is focused onto the sample using a spherical mirror (focal length 5 cm), yielding a spot size of approximately 1 mm. An image of the APT setup can be found in the Appendix B, Fig. B.2. Following evaporation, ions are accelerated toward a position-sensitive detector. Their impact position and time-of-flight are recorded [196], allowing for 3D reconstruction of the specimen. The chemical identity of each ion is determined from its mass-to-charge ratio, while its original position is estimated via back-projection algorithms [197].

The resulting SiO_2 and LaB_6 APT spectra are reported in Fig. 4.4. First, results from silica specimens will be explained, and then the outcomes from the LaB_6 tips.

In APT experiments on needle-shaped silica specimens, evaporation conditions must be carefully controlled due to the insulating nature of the material. For negative THz pulses, the pulse amplitude was thus limited to 0.015 V/nm to provide sufficient transient field enhancement for ion emission while avoiding excessive stress and tip fracture. The DC voltage was automatically adjusted to maintain a constant detection rate of 8% ions per pulse. Clear silicon and silica-related peaks are observed with high signal-to-noise ratio and good mass resolution for the negative pulse (blue spectrum in Fig. 4.4 (a)). These values are comparable to those obtained in UV laser-assisted APT of oxides [198, 199]. The relatively high spectral quality indicates that, under negative THz excitation, the evaporation process is primarily driven by the electric field, with limited thermal contribution: some asymmetry in the peaks suggests that thermal effects are still present but remain secondary. Traces of sodium (Na), hydrogen or other defects are present in a limited quantity.

In contrast, for positive THz pulses (gray in Fig. 4.4 (a)), the amplitude was kept at its maximum while the DC voltage was reduced to maintain a much lower detection rate (0.3% ions

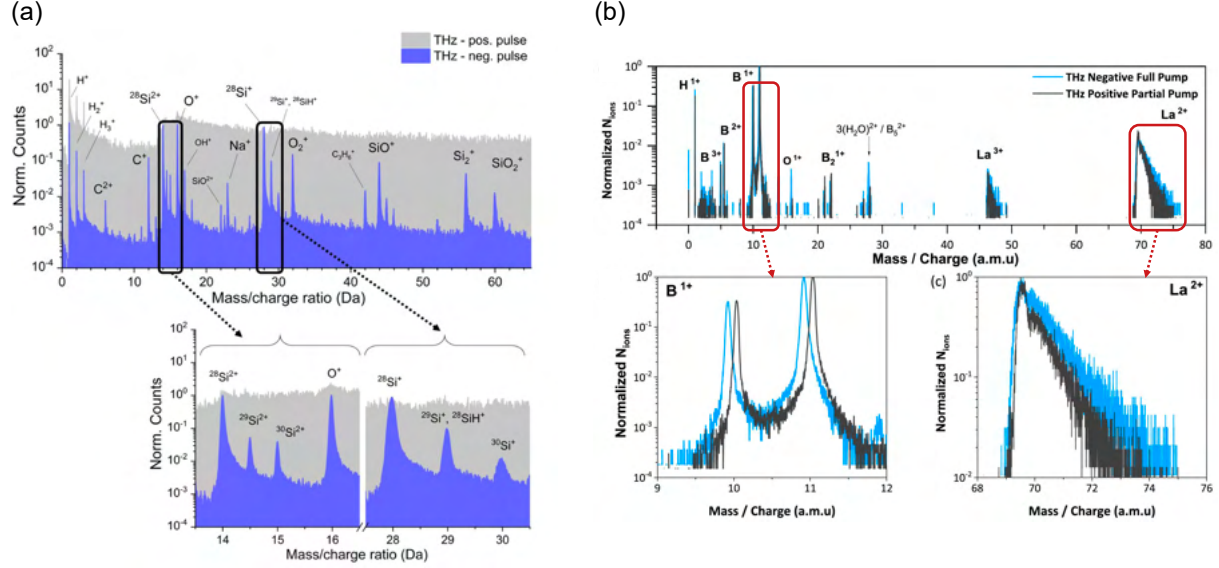


Figure 4.4: (a) Mass spectrum measured using THz-assisted APT on an amorphous SiO₂ sample, comparing THz negative (blue) and positive (grey) polarity pulses. The datasets comprise approximately 4×10^5 ions for negative pulses and 1×10^5 ions for positive pulses, acquired at bias voltages of $V_{DC} = 4$ kV (negative) and 9 kV (positive), with THz field non-enhanced amplitudes of 0.015 and 0.020 V/nm, respectively. The corresponding evaporation rates are 0.08 (negative) and 0.003 (positive) ions per pulse at a temperature of $T = 50$ K. The lower panel shows a zoom-in of the Si²⁺ peak region and its isotopic distribution. (b) Mass spectrum measured using THz-assisted atom probe tomography on a pure LaB₆ sample, comparing negative (blue) and positive (black) THz pulses. The datasets consist of approximately 10^5 ions collected at a bias voltage of $V_{DC} = 6.35$ kV, with a THz field amplitude of 0.02 V/nm for negative pulses and 0.01 V/nm for positive pulses. The evaporation rate is 0.04 ions per pulse at a temperature of $T = 50$ K in both cases. In the lower panels, a zoom of B⁺ and La²⁺ mass peaks is displayed on a semi-logarithmic scale. Adapted from Ref. [184].

per pulse) with respect to the negative polarity. In this regime, the mass spectrum is strongly degraded: only hydrogen-related peaks remain clearly visible, while all other features are masked by a high background. This behavior indicates that evaporation is dominated by thermally assisted processes. The positive THz pulses induce heating at the tip apex, leading to repeated heating and cooling cycles that cause continuous, uncontrolled ion emission and increased noise. That is also why the DC voltage was kept low as mentioned above, because there was already too much uncontrolled thermal evaporation: further increasing the DC field would have led to fractures in the specimen. To conclude, only in negative case the field evaporation process dominates the thermal process, due to the larger THz near field enhancement close to the apex

of the tip.

For LaB₆ samples, the situation differs significantly. The THz amplitude was set to 0.02 V/nm for negative pulses and 0.01 V/nm for positive pulses, while keeping the same DC voltage and a constant detection rate of 4% ions per pulse. The resulting mass spectra in Fig. 4.4 (b) clearly shows a better performance displayed by the positive pulse (black) with respect to the negative (blue). Moreover, the shift in mass-to-charge ratio between positive is attributed to ion acceleration or deceleration induced by the THz field, as also found in similar experiments [200]. Additionally, the peak shapes suggest the coexistence of fast field evaporation and slower thermal processes, as also verified and discussed in [201] for metallic samples behaving similarly. Notably, here negative pulses introduce stronger thermal contributions (longer and greater thermal tails in the blue spectra), whereas positive pulses enhance the fast evaporation component and perform better.

Overall, while silica shows an anomaly, i.e. a strong asymmetry between positive and negative THz excitation –the latter favoring field-driven evaporation–, LaB₆ exhibits an opposite –and more expected– outcome, with the positive polarity effectively inducing evaporation promoted mainly by ion-field and non-thermal mechanisms.

In APT, the evaporation leading to interpretable, highly resolved spectra must be field driven: as the instantaneous total field (DC + enhanced THz) becomes sufficiently large, ion-field evaporation is triggered. In such case, thermal effects such as laser-induced heating play a much smaller role and therefore the experimental signal arrives at high quality in the spectra. Thus, the field enhancement induced by the tip geometry is a pivotal contributor in APT. This means that, the discrepancy in the efficacy between the two polarities could then be explained by differences in such enhancement effect. In fact, as stated in Section 4.1.2, the THz field enhancement promoted by the THz (equation 4.1) is strictly material dependent. Therefore, it is possible that the intrinsic differences between SiO₂ and LaB₆ play a pivotal role in the behavior of the APT field enhancement and the relative total field sensed by the ions while evaporating.

For this reason, it is first essential to visualize how the DC and laser fields (positive, negative polarity) are rearranged within the APT experimental setup. Intuitively, in Fig. 4.5 the DC (black) and THz enhanced field –positive (cyan) or negative (blue) polarity– are represented by arrows indicating their orientation, where here the word orientation refers to the direction in which positive ions are evaporated within the APT chamber: cations always evaporate toward the right side of the chamber in Fig. 4.5, where the detector is placed.

Thus, in the APT experimental setup (Fig. 4.5) the DC field adds to the positive polarity (enhanced) field, but subtracts to the negative one. This means that the evaporation of cations should be more field-dominated in the case of the positive pulse, since the total field should be

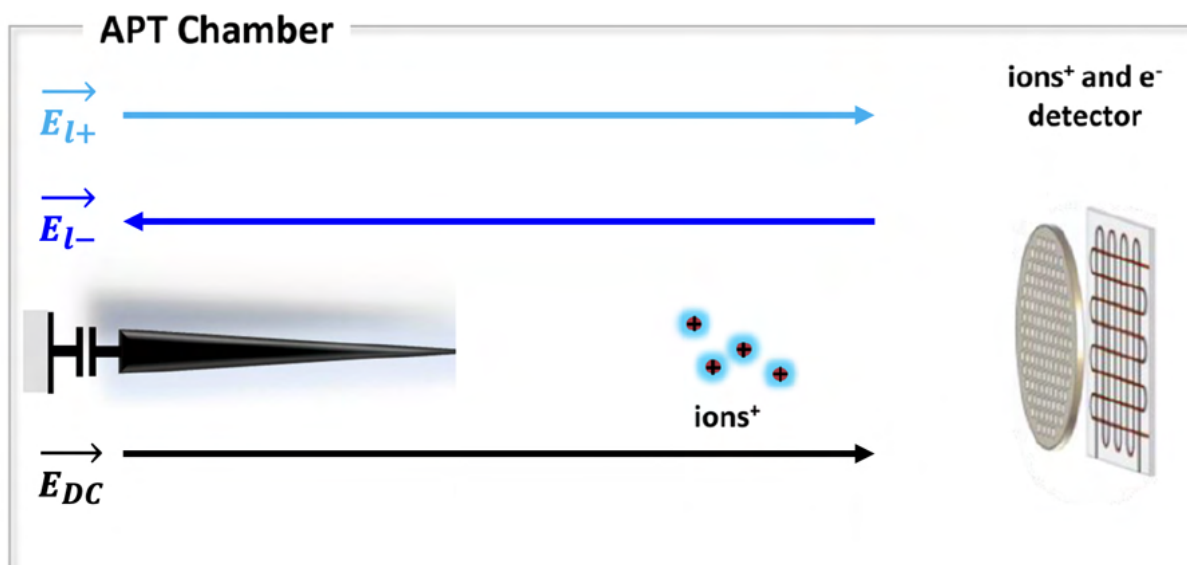


Figure 4.5: Scheme of the orientation of each field within the APT Chamber. The static voltage is represented in black, whereas the positive and negative polarity lasers in cyan and blue, respectively. Cations are evaporated towards the right side of the chamber. Notably, the DC field adds to the positive pulse, but subtracts to the negative one.

bigger, and such molecular ions should be strongly induced to leave the sample, reaching the detector with higher efficiency. As expected, such scenario is confirmed in fact for metallic and semi-metallic tips like LaB_6 , but strangely not for silica samples. Therefore, it is evident that in the case of SiO_2 the THz pulse polarity has an unknown effect on the two main factors that typically promote field-driven evaporation of cations: THz field enhancement and total field summation (DC + enhanced THz).

4.3 The *ab initio* MD Study

Finding an answer to the discrepancy found in the experimental results of the last section is relevant, as it could also lead to new knowledge about insulating and semi-metallic in the context of APT, reaching new frontiers of applications. As stated in Chapter 1, Section 1.7.2, *in silico* techniques can provide some intuition and insights in order to formulate a meaningful hypothesis. In fact, a time-dependent framework like *ab initio* Molecular Dynamics simulations is essential to capture the simultaneous interaction of the sample with THz and static fields. Specifically, Time-dependent density functional theory (TDDFT) [98] provides a rigorous quantum-mechanical description of electronic excitations in external fields and can be naturally coupled to MD workflows within the Ehrenfest formalism [99].

The evaporation of cations from a solid surface is a complex multiscale process involving several coupled mechanisms. While *ab initio* approaches provide a rigorous framework to investigate it, a fully realistic description remains out of reach. This is mainly due to the coexistence of electronic ionization and thermal effects during APT experiments. Ionization can be treated within first-principles methods, whereas thermal processes –such as electron-phonon coupling and lattice heating– occur over longer and often stochastic timescales, making them difficult to capture accurately within TDDFT and Ehrenfest frameworks. In particular, TDDFT, being a mean-field approach, does not fully account for decoherence or dissipative effects realistically. For this reason, the present work focuses on field-induced evaporation mediated by ionization, which is both well defined and directly accessible within an *ab initio* time-dependent framework. A detailed discussion about the *ab initio* Molecular Dynamics procedure employed in this thesis (TDDFT and Ehrenfest integrations schemes) is provided in Chapter 2, Sections 2.4 and 2.5).

Typically, the TDDFT formulation has high computational costs, often needed to achieve the accuracy required to describe the APT evaporation processes. Such aspect of TDDFT limits significantly the size of the simulated system. For this reason, to describe and investigate the behavior of a silicon dioxide matrix under an electrostatic field and THz laser irradiation, a much simpler model bound by TDDFT limitations was utilized, consisting of only one $\text{Si}(\text{OH})_4$ molecule (the matrix precursor). Such model will be named the First Model, and its aim is to provide a computational procedure to model THz-mediated ion field evaporation, with a particular focus on the correct reproduction of the THz pulse polarity, a relevant parameter to consider in the context of ion emission and APT. Fig. 4.6 represents intuitively its characteristics, that can be associated to the APT setup in Fig. 4.5.

However, a system composed by a single molecule is quite limited by the lack of a band structure and the size constraints of a silica matrix. In fact, electrons are either confined or free to move in a realistic scenario, according to the specimen properties (insulating or metallic material). For this reason, a second simplified toy model is introduced (referenced as the Second

Model), with the purpose (i) to bridge the gap between the first $\text{Si}(\text{OH})_4$ model (molecules) and silica matrices (bulks) and (ii) to account for the diverse electron mobility in insulating SiO_2 specimens compared to semi metallic LaB_6 samples. In this manner, the Second Model provides a possible explanation to the found experimental results, by relating the chemical nature of the analyzed samples with the THz pulse polarity.

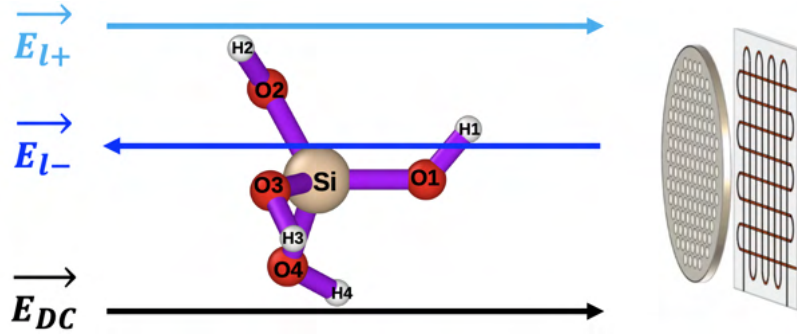


Figure 4.6: The First Model: a $\text{Si}(\text{OH})_4$ molecule is irradiated by a THz and a DC field. The pulse can have positive (cyan) or negative (blue) polarity. More details can be found in the Section 4.4.

4.4 Simulation Protocol

This section describes the simulation protocol utilized for the First Model. The procedure employed for the Second Model shares almost every detail with the First model workflow, and the small but relevant differences will be directly introduced in Section 4.5.2.

4.4.1 Software Details

All calculations were performed using the OCTOPUS code [202, 203]. OCTOPUS is based on a real-space grid representation, where wave functions are discretized directly in real space rather than expanded in a basis set. This approach allows for an efficient and flexible treatment of time-dependent problems and external fields without basis-set limitations.

4.4.2 External Field

The system is subjected to an external field that can be static (E_{DC}), time-dependent (E_{laser}), or both (like in the APT setup). The total external potential is written as:

$$v_{\text{ext-apt}}(x, t) = x \cdot \hat{e} \left[E_{\text{DC}} + E_{\text{laser}} \exp \left(-\frac{(t - t_0)^2}{2\sigma^2} \right) \sin(\omega t) \right], \quad (4.3)$$

where $\hat{e}$ is the polarization vector, $\omega/2\pi = 5$ THz, $t_0 = 50$ fs, and $\sigma = 14.13$ fs. The static field E_{DC} is always positive, while E_{laser} can be either positive or negative, corresponding to the two THz pulse polarities.

The laser field is modeled as a Gaussian envelope multiplied by a sinusoidal carrier, allowing flexible tuning of frequency, duration, and intensity. The dipole approximation is valid in our simulations, as the simulation box size ($a = 2$ nm) is much smaller than the THz wavelength ($\lambda \sim 6 \times 10^4$ nm); therefore, it is a classical field (so, not quantised). The field mimics the effect of a single-cycle THz pulse, as reported in Fig. 4.7, lasting 0.1 ps (due to TDDFT time constraints), showing both the positive (cyan) and the negative (blue) pulse polarity cases. Notably, the red curve in Fig. 4.7 refers to the UV laser, where polarity does not really play a relevant role here (too fast oscillations in the picosecond timescale).

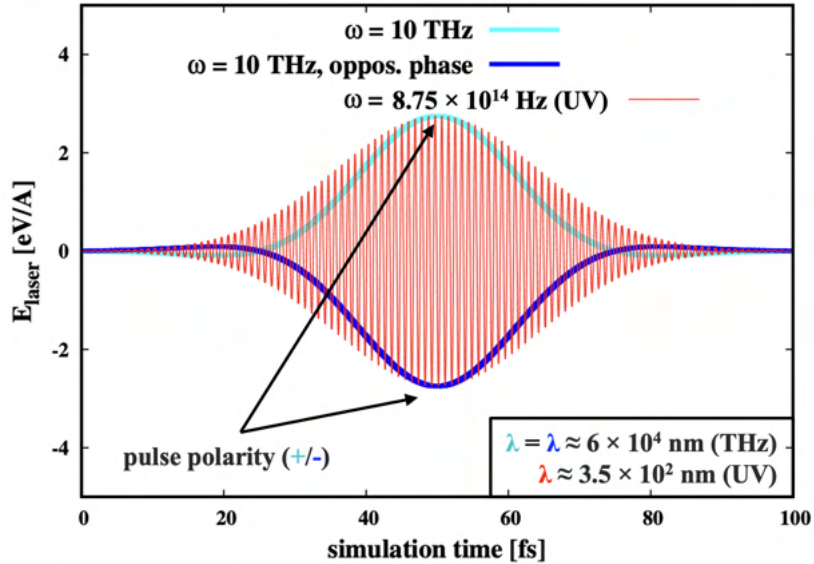
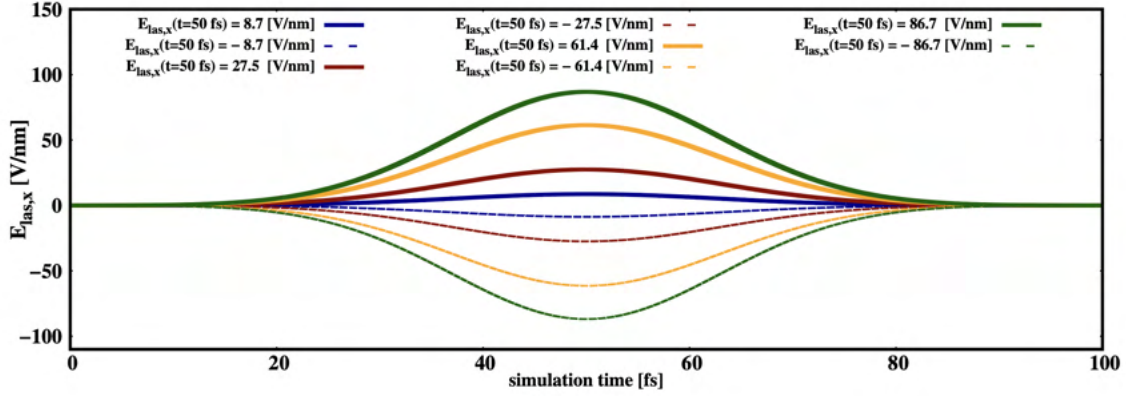


Figure 4.7: Model of a Single-Cycle THz Pulse of positive (cyan) or negative (blue) polarity. For comparison purposes, a UV pulse is shown in red.

The laser frequency utilized is 10 THz (1–10 THz were utilized in the experimental workflow). A range of THz field amplitudes was considered: seven values for each polarity, spanning from 8.75 to 86 V/nm, resulting in a total of 14 field configurations, as can be seen by Fig. 4.8. These values are in the order of the V/nm, as they already reproduce the antenna effect, i.e. the enhancement of field near the tip apex (as in experiments).

The static electric field values used are $E_{\text{DC}} = (0,0,0)$, $(5,0,0)$, $(10,0,0)$ V/nm. These values ensure that electrons remain bound in the absence of the laser field. An explicit treatment of the electric field is essential to prevent the electrons from traveling to the edges of the box

(a)



(b)

$$\begin{aligned}
 \xrightarrow{E_{dc}} & \begin{cases} (0, 0, 0) \frac{v}{nm} \\ (5, 0, 0) \frac{v}{nm} \\ (10, 0, 0) \frac{v}{nm} \end{cases} \xrightarrow{E_{l+}} \begin{cases} (8.75, 0, 0) \frac{v}{nm} & (27.5, 0, 0) \frac{v}{nm} \\ (41.1, 0, 0) \frac{v}{nm} & (48.8, 0, 0) \frac{v}{nm} \\ (54.5, 0, 0) \frac{v}{nm} & (61.4, 0, 0) \frac{v}{nm} \\ (86.7, 0, 0) \frac{v}{nm} & \end{cases} \xrightarrow{E_{l-}} \begin{cases} (-8.75, 0, 0) \frac{v}{nm} & (-27.5, 0, 0) \frac{v}{nm} \\ (-41.1, 0, 0) \frac{v}{nm} & (-48.8, 0, 0) \frac{v}{nm} \\ (-54.5, 0, 0) \frac{v}{nm} & (-61.4, 0, 0) \frac{v}{nm} \\ (-86.7, 0, 0) \frac{v}{nm} & \end{cases}
 \end{aligned}$$

Figure 4.8: (a) Values of THz Laser Amplitudes (b) coupled with DC Field ones. For E_{laser} , seven positive and seven negative polarity pulses were employed, combined with three static DC fields. The values are reported as the modulus of the field, i.e. the maximum value of the peak (found at $t = 50$ fs).

at a large field amplitude. While modern polarization theory can address this problem, an explicit treatment of E_{DC} is retained, as it has proven useful in previous studies to describe the evaporation dynamics of atoms from molecules and periodic systems [204, 205, 206, 207].

Importantly, the static field is time-independent but spatially varying, enabling the emulation of THz field enhancement in the vicinity of the tip apex. As shown in Fig. 4.9, the irradiated molecule is positioned at the extremity of the needle-shaped APT specimen, in close proximity to the tip (see points (a) and (b)). Owing to its small size, the molecule effectively probes the maximum field value; indeed, for positions near $x \approx 0$, the field E_{DC} is approximately uniform.

This component of the model is therefore primarily intended to demonstrate that field enhancement can be represented, even though the physical phenomenon occurs over larger length scales (on the order of tens of nanometers). In most cases, the E_{DC} term in equation 4.3 remains effectively constant and close to the three values introduced earlier.

However, the analytical form of the green curve in Fig. 4.9 (b) mimicking the position-dependent potential can be expressed as a smoothly switched-on linear ramp with a constant offset. It is

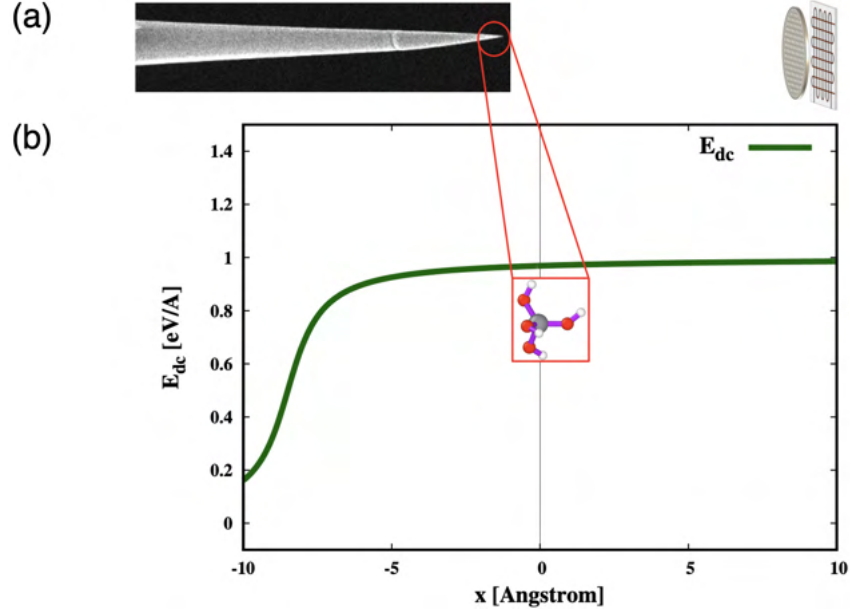


Figure 4.9: Modelization of the static electric field E_{DC} . (a) Sketch of the APT experimental setup. (b) Plot of E_{DC} in function of the x-dimension of the simulation box, mimicking the field enhancement effect.

defined by:

$$V_{DC}(x) = 0.5 \frac{\text{eV}}{\text{\AA}} \left(0.5 + \frac{1}{\pi} \arctan(1.2(x + 8.5 \text{\AA})) \right) (x + 8.5 \text{\AA}) + |V_0|, \quad (4.4)$$

where the constant offset V_0 is given by evaluating the same expression at a reference point:

$$V_0 = 0.5 \frac{\text{eV}}{\text{\AA}} \left(0.5 + \frac{1}{\pi} \arctan(1.2(-1.0 \text{\AA})) \right) (-1.0 \text{\AA}). \quad (4.5)$$

The factor

$$S(x) = 0.5 + \frac{1}{\pi} \arctan(1.2(x + 8.5 \text{\AA})) \quad (4.6)$$

acts as a smooth step function, interpolating from 0 to 1 around $x = -8.5 \text{\AA}$. Thus, the full potential $V(x)$ represents a linear ramp (constant field) that turns on smoothly, avoiding discontinuities in the simulation box, with a slope of $0.5 \text{ eV/\AA}$ and a constant energy shift for reference.

4.4.3 Time Propagation Scheme

The coupled electron–ion dynamics is described within the Ehrenfest approximation (atomic units):

$$i\frac{\partial\psi_j(\mathbf{r},t)}{\partial t} = \left[-\frac{\nabla^2}{2} + v_{\text{ext-apt}}(\mathbf{r},t) + v_{\text{H,xc}}(\mathbf{r},t) + \hat{W}_{en} \right] \psi_j(\mathbf{r},t), \quad (4.7)$$

$$M_J\frac{\partial^2\mathbf{R}_J}{\partial t^2} = -\nabla_{\mathbf{R}_J} \left[v_{\text{ext-apt}}(\mathbf{R}_J,t) + \hat{W}_{nn} + \int d^3r n(\mathbf{r},t)\hat{W}_{en} \right], \quad (4.8)$$

where ψ_j is the electronic wave function of the occupied orbitals, $v_{\text{ext-apt}}$ is the external scalar potential acting on both electrons at position $\{\mathbf{r}\}$ and nuclei at position $\{\mathbf{R}\}$, $v_{\text{H,xc}}$ is the Hartree plus exchange–correlation potential, $\hat{W}_{nn}$ is the Coulomb interaction between the nuclei, M_J is the mass of the J -th nucleus, $\hat{W}_{en}$ is the Coulomb interaction between the electrons and the nuclei, and $n(\mathbf{r},t) \equiv \sum_j^{\text{occ}} |\psi_j(\mathbf{r},t)|^2$ is the time-dependent electron density. This represents a variation of the coupled scheme presented in the self-consistent equations 2.71, 2.72 and 2.88, to which the external field in equation 4.3 was added. In Chapter 2, Section 2.5 there is an exhaustive explanation of this TDDFT + Ehrenfest scheme. Additional implementation details are provided in Ref. [192], which also includes Refs. [208, 209, 210]. Electron–ion interactions are described using pseudopotentials within the local density approximation (LDA) [208], and the adiabatic LDA is employed for time-dependent simulations.

Here, the OCTOPUS implementation of the Approximated Enforced Time-Reversal Symmetry (AETRS) propagator was employed. AETRS is closely related to the Crank–Nicolson (CN) propagators (see Section 2.5), as both approximate the exact time-evolution operator using the Hamiltonian evaluated at the midpoint of the time step. AETRS applies the exponential form $\exp[-iH(t + \frac{\Delta t}{2})\Delta t]$, while Crank–Nicolson uses a rational (Padé) approximation of the same exponential, $(1 + \frac{i\Delta t}{2}H)^{-1}(1 - \frac{i\Delta t}{2}H)$. As a result, both methods are unitary, time-reversal symmetric, and second-order accurate, differing mainly in that AETRS uses the exponential operator directly, whereas Crank–Nicolson replaces it with a numerically convenient rational form.

4.4.4 Si(OH)₄ Atomic Coordinates and Simulation Box Parameters

The equilibrium coordinates of Si(OH)₄ were taken from previous results, generated with DFT procedures (see Chapter 3, Section 3.4.2). The molecule has a tetrahedral shape with the silicon atom coordinated by four hydroxyl groups, and has 32 valence electrons. The simulation box employed for the calculation is a 10 Å³ cubic box, chosen to conserve the system’s energy without time-dependent fields. The lattice spacing is 0.16 Å in each dimension, ensuring total energy convergence below 5 meV.

4.4.5 Optimization under Static Electric Fields

The DFT equilibrium structure of the $\text{Si}(\text{OH})_4$ molecule was modeled under different values of the external static field and is used as the initial configuration for the temporal evolution of the system. To do so, an optimization strategy was employed, using the FIRE algorithm [210] with a force threshold of $0.03 \text{ eV/\AA}$. To prevent global translations and rotations of the molecule under the static external field, we clamped the central Si atom, the y and z coordinates of the oxygen atom O1, and the y coordinate of the hydrogen atom H1 (see Fig. 4.10 for the atom names) during the optimization.

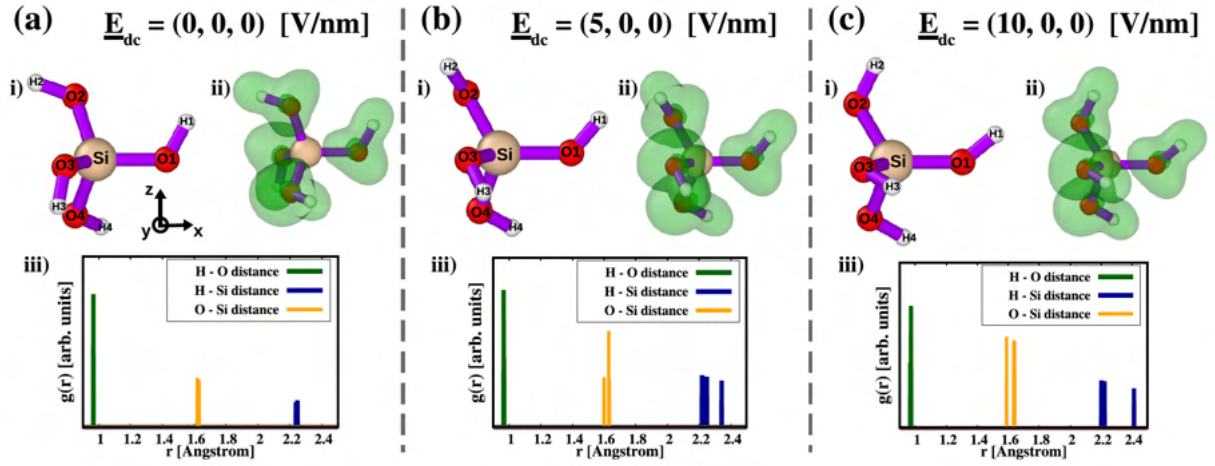


Figure 4.10: $\text{Si}(\text{OH})_4$ molecule as a function of the static external field: (a) $E_{\text{DC}} = (0, 0, 0)$; (b) $(5, 0, 0)$; and (c) $(10, 0, 0)$ V/nm. The insets (i) at the top left of panels (a)–(c) show the optimized geometries, while the insets (ii) at the top right display the electron localization function (ELF) at an isosurface value of 0.66. The bottom insets (iii) report the corresponding pair distribution functions. The geometries were optimized by clamping the Si atom, the y and z coordinates of the oxygen atom O1, and the y coordinate of the corresponding hydrogen atom H1. Adapted from Ref. [184].

4.4.6 Boundary Conditions

Absorbing boundary conditions were applied to avoid artificial reflections of emitted electrons. This is implemented *via* a mask function that smoothly damps the electronic density near the box edges:

$$M(\mathbf{r}) = \prod_{i=x,y,z} m_i(r_i) \quad (4.9)$$

where the one-dimensional mask along direction i is given by:

$$m_i(r_i) = \begin{cases} 1, & |r_i| \leq R_i - L_i \\ \cos^p \left[\frac{\pi}{2} \frac{|r_i| - (R_i - L_i)}{L_i} \right], & R_i - L_i < |r_i| < R_i \\ 0, & |r_i| \geq R_i \end{cases} \quad (4.10)$$

where r_i is the coordinate along the i -th Cartesian direction, R_i is the half-length of the simulation box along that direction, and L_i defines the width of the absorbing region near the box boundaries. The parameter p controls the smoothness of the mask decay, with typical values $p = 1$ or $p = 2$. The function is equal to unity in the central region ($|r_i| \leq R_i - L_i$), smoothly decreases within the region of the absorbing layer ($R_i - L_i < |r_i| < R_i$), and vanishes at the box edges ($|r_i| \geq R_i$). Notably, each m_i transitions smoothly from 1 to 0 within an absorbing region.

4.4.7 Production Run Details

The time step was 4×10^{-4} fs, and simulations were run for 100 fs (approximately 2.5×10^5 steps). A total of 42 simulations were performed, combining three values of E_{DC} (0, 5, 10 V/nm) with the 14 THz field configurations (seven positive and seven negative polarities). During the dynamics, the positions of the silicon atom and the O2H2, O3H3 and O4H4 (labels follow the notation of Figs. 4.6 and 4.10) ions are fixed to prevent the molecule from drifting. In fact, the main purpose of the simulation is to study the evaporation of the O1 and H1 atoms in form of ions, in response to static and THz fields, which is similar to the evaporation of single atoms from a solid surface [100].

4.5 Results of *ab initio* MD Simulations

This section reports the results from the *ab initio* MD simulations. First, the outcomes of the First Model are described, in order to get a good assessment of the modeling of a THz laser source coupled with a static DC field on a simpler system, the $\text{Si}(\text{OH})_4$ molecule. Secondly, the findings obtained from the Second Model will be introduced, with the purpose to complete what found for the First Model, trying to get closer to the real experimental setup of APT (a needle-shaped silica matrix). Details about the created models can be found in Sections 4.3.

4.5.1 Modeling the THz-assisted Field Evaporation Mechanism in Molecules: The First Model

A Brief Summary of the Key Modeling Aspects

To interpret the puzzling APT experimental findings concerning semi-metallic and insulating samples in the context of single-cycle THz pulse polarity, a detailed first-principles theoretical analysis of atom field evaporation under combined static (DC) and time-dependent (THz) fields was conducted. The microscopic mechanism of field evaporation was explored by directly coupling RT-TDDFT –essential to describe real-time electron dynamics— with Ehrenfest dynamics – an effective approximation for nuclear motion. This coupled TDDFT+Ehrenfest framework has been used previously to describe coupled electron–ion motion in atomic environments and surface processes [211, 212].

As already stated in Section 4.3, the First Model uses $\text{Si}(\text{OH})_4$ as a molecular surrogate for key evaporation physics. Despite being completely aware that molecular and bulk structural and electronic properties differ significantly, TDDFT presents computational constraints, not allowing to emulate *in silico* the real-time electron dynamics of bulks. This limitation and its implications will be discussed further with reference to an additional simplified “toy” model, the Second Model (see Section 4.5.2). Throughout the time-dependent simulations, the **positions of the Si atom and selected O/H groups (O_2H_2 , O_3H_3 , O_4H_4) are fixed**, so that only the relevant functional groups (O_1 , H_1) can respond to the external fields. This step was essential because in this way it is possible to know which molecular ion is evaporated. Notably, the electrons of all fixed atoms are allowed to follow their natural dynamics within this APT context. The labeling of atoms follows Fig. 4.6. For this reason, the focus of the work is on the evaporation behavior of O_1 and H_1 atoms.

Three static field strengths were considered: $E_{\text{DC},x} = 0$ V/nm, $E_{\text{DC},x} = 5$ V/nm and 10 V/nm, using the molecular geometries shown in Figs. 4.10 (a), (b) and (c) as starting configurations. These configurations do not represent the full configurational space –many more

angles between the external fields and the molecule are possible. In fact, the O₁–Si bond was specifically aligned along the x-axis so that the contributions from the Coulomb force, static field, and THz pulse can be uniquely assessed. Consequently, the present results serve as an upper bound on the critical fields necessary for ion evaporation; in realistic bulk silica, additional factors such as matrix defects and angular distributions would further complicate the picture. Notably, the evaporation condition will be thus toward the right side, positive values of the x-axis, where the detector is virtually placed in APT experiments. A full comparison between the experimental and the computational setups can be obtained by comparing Figs. 4.5 and 4.6. A total of 42 simulations were performed, combining three values of E_{DC} with the 14 THz field configurations (seven positive and seven negative polarities). Notably, every time that a THz field amplitude/intensity is cited, like for example 50 V/nm, such number corresponds to its value at $t \approx 50$ fs, near the peak of the THz pulse (see Fig. 4.8 (a)).

Hydrogen and Oxygen Evaporation

A distinctly different evaporation behavior was found depending on both the peak amplitude and the polarity of the THz pulses. The time-dependent results are collected in Fig. 4.11, where the H₁–Si distance is shown in (a), (c) and (e), whereas the O₁–Si distance in (b), (d) and (f). Different symbols represent different THz intensities and static field baselines; positive THz pulses are shown in (a)–(d) and negative pulses in (e)–(f).

Under positive THz fields, hydrogen evaporation is strongly influenced by both the magnitude of the static field and the peak amplitude of the THz pulse. For instance, when the DC field is absent ($E_{DC,x} = 0$ V/nm), even moderate THz fields such as 27.5 V/nm fail to remove H₁ from the molecule (small dots in Fig. 4.11 (a)). Conversely, the addition of a static field of 10 V/nm reduces the threshold for hydrogen evaporation dramatically, allowing H₁ to be released at the same THz amplitude (small dots in Fig. 4.11 (c)). This effect highlights the synergistic interplay between static and THz fields, which combines to produce a larger effective field along the molecular axis. Once evaporation begins, H₁ is observed to rapidly move away from the Si atom, with the H₁–Si distance exceeding 10 Å in several high-field cases, signaling complete detachment. The rapidity of the hydrogen escape reflects its low mass and the strong Coulomb repulsion experienced once electrons are extracted out of the molecule. Furthermore, the charge of H₁ remains consistently +1, indicating that it leaves the molecule fully ionized regardless of field polarity or magnitude, a behavior consistent with the expected proton-like nature of the evaporating hydrogen in strong fields (see Table 4.1).

Intriguingly, for negative pulses, a much higher value of -54.5 V/nm for the THz laser under the effect of a 10 V/nm electric field is needed to obtain the evaporation of one proton, as seen from Fig. 4.11 (e). Apparently, it seems that the positive pulse performs better for the hydrogen

atom.

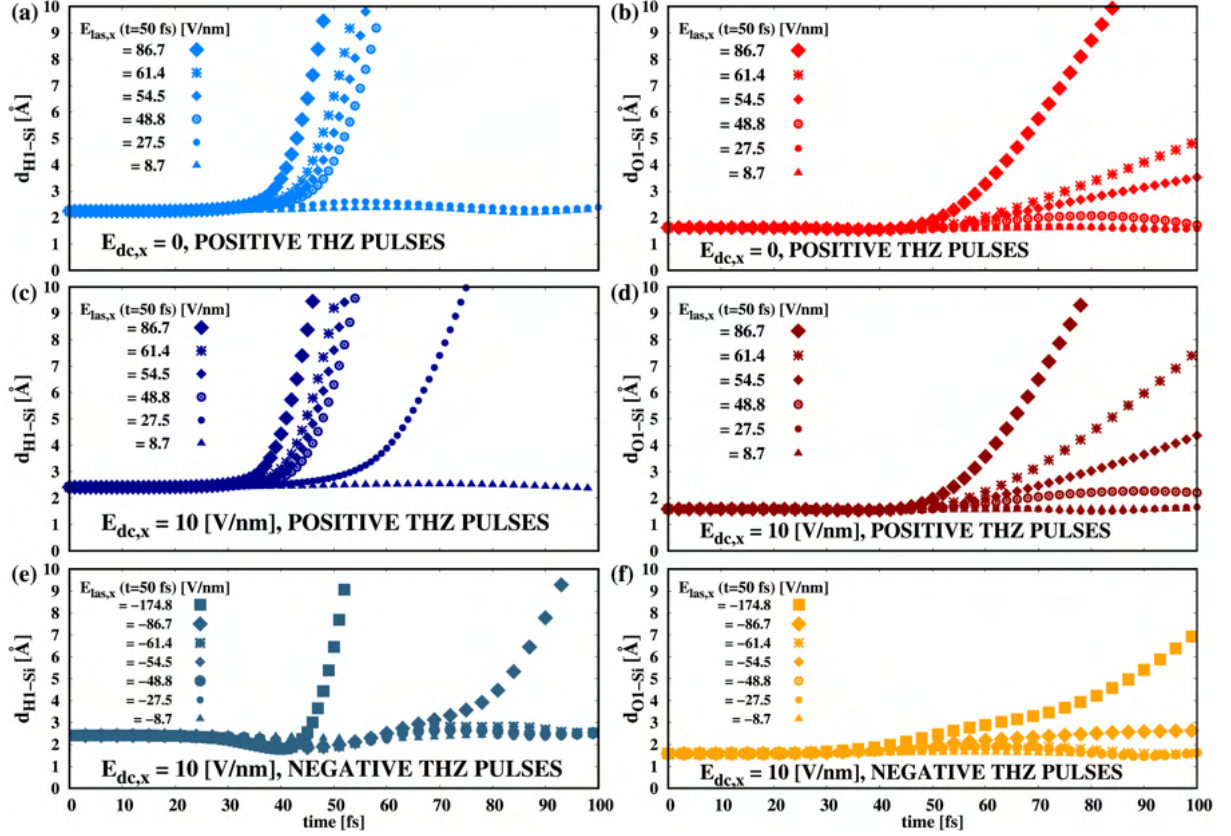


Figure 4.11: Time evolution of atomic distances under THz-assisted fields of different pulse polarities. Panels (a), (c), and (e) show the distances between the hydrogen atom H_1 and the silicon atom as a function of time. Panels (b), (d), and (f) show the distances between the oxygen atom O_1 and the silicon atom as a function of time. The labeling of atoms follows the convention introduced in Fig. 2. Panels (a)–(d) correspond to positive THz pulses, while panels (e) and (f) correspond to negative THz pulses. Adapted from Ref. [184].

Oxygen evaporation exhibits a more complex dynamics due to the stronger Si–O bond and the higher mass of the atom. A clear threshold effect is observed: only THz fields above 54.5 V/nm (positive polarity) can break the Si– O_1 bond (red diamonds in Figs. 4.11 (b) and (d)). Below this threshold, the O_1 atom remains attached, even if H_1 evaporates. The O_1 –Si distance increases more slowly than H_1 –Si, reflecting both the bond strength and the greater inertia of oxygen. This slow evolution complicates the precise determination of the evaporation point within the simulation time window.

In the negative polarity scenario, evaporation thresholds are significantly higher (Figs. 4.11

(e) and (f)), about three times larger than that for positive pulses (an enormous value of -174.8 V/nm is needed to trigger the mechanism), suggesting a bad evaporation efficiency.

Pulse Polarity, Ionization and Electron Loss Govern the Field-driven Evaporation Mechanism

Interestingly, there is a qualitative disparity between positive and negative polarity. Such difference mirrors experimental findings in semi-metallic samples such as LaB₆, where negative pulses are less effective than positive ones. In contrast, the same polarity dependence is not found experimentally in insulating silica matrices. This polarity effect arises because the relative alignment of static and oscillating fields governs the effective net field experienced by the electrons. When the THz and DC fields share the same sign, the total field strength is enhanced, promoting stronger ionization —the driver of bond breaking and atomic evaporation. When they oppose each other, partial cancellation diminishes the total field and raises the threshold for ion evaporation. The next paragraph focuses on a more detailed description of such process.

Ehrenfest dynamics nonetheless yields rich information on forces, velocities, and kinetic energy of the escaping ions. This data, taken together with charge density analyses, highlights the critical role of ionization: intense THz pulses promote electron loss and a concomitant accumulation of positive charge, as visualized in Fig. 4.12. Here, electronic density (green) and evaporating ions (O (red), H (white)) are shown at different simulation times for ± 61.4 V/nm peak THz amplitudes. A stark difference emerges at the height of the pulse ($t_{\text{sim}} \approx 50$ fs): under positive pulses, electron depletion occurs toward the left edge of the simulation cell, whereas under negative pulses electrons migrate toward the right, where the detector is virtually placed (see Fig. 4.6). This anisotropic electron depletion drives an evolving net positive charge on the molecule, which in turn accelerates the departing ions. For positive pulses, H₁ and O₁ experience strong Coulomb repulsion from the fixed molecular scaffold after about 48 fs, enabling their release even without the static field. In contrast, under negative polarity, the effective field is weaker due to cancellation, producing lower ionization and requiring much higher THz intensities to trigger evaporation.

In summary, the polarity of the THz pulse plays a critical role, as it maneuvers both ionization and the subsequent electron loss, key aspects for the evaporation process. In fact, the simulations reveal that intense THz fields induce electron flows out of the molecule, creating a net positive charge on the fixed part of Si(OH)₄. If such loss is large enough, the subsequent accumulation of positive charge generates strong Coulomb forces on the evaporating ions, accelerating them away from the molecular scaffold, as illustrated in Fig. 4.12.

Table 4.1 summarizes the final charge states and distances of the relevant atoms, providing a complete quantitative picture. In fact, the emitted charge distribution also provides insight into

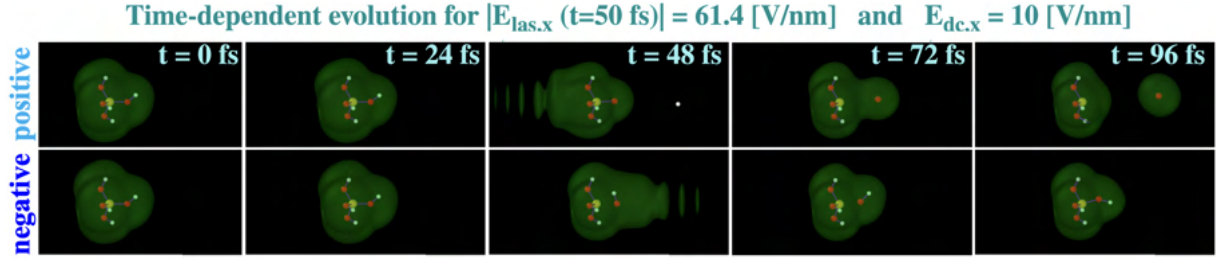


Figure 4.12: Isosurface plots of the electronic density (10^{-4} a.u.) and ionic positions at five different time snapshots during the temporal evolution for $E_{\text{laser},x}(t = 50 \text{ fs}) = 61.4 \text{ V/nm}$. The upper panels correspond to the evolution under a positive THz pulse, while the lower panels show the evolution under a negative THz pulse. As noted by the images, the loss of electrons occurs in the two different directions (left - positive, right - negative) along the x-axis. Adapted from Ref. [184].

the total number of electrons removed. For each simulation, the sum $N_{\text{mol}}^* + N_{\text{H1}}^* + N_{\text{O1}}^*$ gives the number of electrons emitted. Hydrogen always contributes +1 electron when evaporated, while oxygen contributes depending on the THz amplitude and polarity: its charge increases with the laser intensity only for positive pulses, whereas under negative ones, it does not detach within accessible field ranges. The final charge on oxygen increases with the THz amplitude, but only for positive pulses. In contrast, for negative pulses, the O_1 atom often remains attached or neutral, further underscoring the directionality dependence of the field-assisted evaporation. These results suggest that the evaporation mechanism is highly nonlinear, as small changes in field strength or polarity produce qualitatively different outcomes in atom detachment and ionization.

Quick Summary

In summary, *ab initio* MD simulations indicate that:

1. The presence of a DC field lowers the evaporation thresholds.
2. Hydrogen atoms are easily evaporated under combined DC and THz fields.
3. Oxygen atoms require larger THz amplitudes, and their evaporation is highly sensitive to pulse polarity.
4. Negative THz pulses are less effective due to partial cancellation with the DC field, requiring higher amplitudes to achieve evaporation with respect to positive THz pulses, systematically inducing evaporation at lower amplitude values.

Table 4.1: Data from the Ehrenfest dynamics simulations. $d_{\text{H1-Si}}$ and $d_{\text{O1-Si}}$ represent the distances between the hydrogen atom H1 and the Si atom, and between the oxygen atom O1 and the Si atom, respectively. N_{mol}^* denotes the final total charge of the fixed molecule, while N_{H1}^* and N_{O1}^* indicate the charges of the evaporated hydrogen and oxygen atoms. The latter two values are reported only if the respective atom-Si distance exceeds 3 Å. The total number of electrons emitted is given by the sum $N_{\text{mol}}^* + N_{\text{H1}}^* + N_{\text{O1}}^*$. The symbol “-” indicates a neutral state.

$E_{\text{DC},x}$ (V/nm)	$E_{\text{las},x}$ (V/nm)	N_{mol}^*	N_{H1}^*	N_{O1}^*	$d_{\text{H1-Si}}$ (Å)	$d_{\text{O1-Si}}$ (Å)
0	8.7	-	-	-	2.30	1.61
10	8.7	0.1	-	-	2.36	1.59
10	-8.7	-	-	-	2.48	1.59
0	27.5	0.8	-	-	2.39	1.58
10	27.5	0.6	1.0	-	>10	1.58
10	-27.5	0.7	-	-	2.41	1.65
0	41.1	1.2	1.0	-	>10	1.60
10	41.1	2.0	1.0	-	>10	1.49
10	-41.1	1.1	-	-	2.60	1.64
0	48.8	1.6	1.0	-	>10	1.71
10	48.8	2.6	1.0	-	>10	2.20
10	-48.8	2.0	-	-	2.55	1.64
0	54.5	2.4	1.0	0.5	>10	3.52
10	54.5	2.5	1.0	0.9	>10	4.37
10	-54.5	2.4	-	-	2.58	1.62
0	61.4	2.7	1.0	0.7	>10	4.88
10	61.4	2.8	1.0	1.0	>10	6.27
10	-61.4	1.9	-	-	2.55	1.59
0	86.7	3.9	1.0	1.9	>10	>10
10	86.7	4.0	1.0	2.1	>10	>10
10	-86.7	4.2	1.0	-	>10	2.64
10	-174.8	9.8	1.0	1.5	>10	7.11

5. The pulse polarity plays a central role, by promoting in different ways ionization and subsequent electron loss from the molecule, generating Coulomb forces that drive atomic ejection.

Taken together, these findings provide a quantitative picture of atomic evaporation in a molecule under THz-assisted conditions, offering mechanistic insights into the nonlinear interplay of field amplitude, pulse polarity, ionization and atomic mass. Despite the isolated-molecule model, the Fist Model captured the essential physics of field-driven evaporation process. The clear asymmetry with respect to the pulse polarity variable evinced a striking fact: molecules behave consistently with experimental observations in semi-metallic systems such as LaB_6 , where negative pulses exhibit a lower efficiency in inducing evaporation than their positive counterparts.

4.5.2 Bridging the Gaps between Bulks and Molecules: The Second Model

The theoretical outcomes found in the last section depart significantly from APT experiments on SiO_2 matrices: simulations suggest lower thresholds for positive THz pulses than negative ones, contrary to the experimental mass-spectra results for amorphous silica, reported in Fig. 4.4. This discrepancy likely stems from the use of an isolated $\text{Si}(\text{OH})_4$ molecule in vacuum as a model. In a bulk system, excited electrons have a finite lifetime and can be extracted if the external field energy exceeds the band gap + work function (insulators) or just the work function (metals), when states at the conduction band bottom are accessible. The subsequent electron dynamics –whether they remain bound or are transported– crucially distinguishes insulating versus metallic/semi-metallic behavior, and consequently the atom evaporation mechanism.

For these reasons, the Second Model has been developed. This model aims to provide a qualitative interpretation of the discrepancies observed between theoretical predictions and experimental results, particularly in relation to the polarity of THz pulses and their influence on the electron redistribution under external perturbations, central to understanding these discrepancies.

Features of the Second Model

The Second Model introduces a simplified and fictitious representation of a bulk material. This model is designed to mimic the essential features of a needle-shaped specimen typically used in an APT setup, while remaining computationally tractable. To achieve this, the simplest possible bound system has been considered, namely a hydrogen molecule (H_2). The same framework described in Section 4.4 was applied, based on TDDFT+Ehrenfest dynamics. However, for simplicity, only one dimension (1D) was considered, with a box size of 80 Å, a lattice spacing of 0.05 Å, a time step of 5×10^{-4} fs for the time integration of the equations of motion, and a

total of 3.2×10^5 steps to follow the dynamics of the system even after removal of the THz field. The reduction to 1D should significantly decrease the computational cost and complexity, while still retaining the fundamental physical mechanisms governing electron–nucleus interactions and field-induced dynamics.

The full configuration of the H_2 system is illustrated in Fig. 4.13. It is important to emphasize that this model is not intended to provide quantitatively accurate predictions of field strengths or evaporation rates, but rather to capture and interpret the underlying physical mechanisms in a controlled and transparent manner.

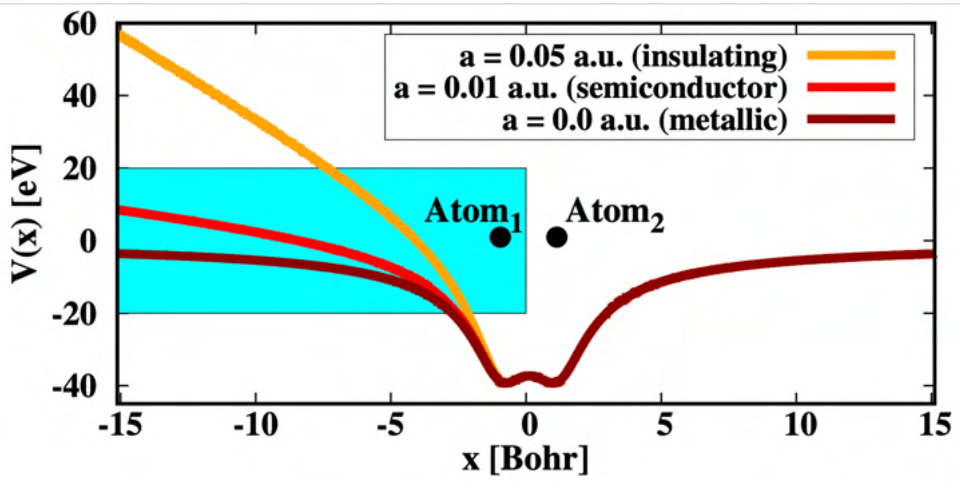


Figure 4.13: The one-dimensional potential of the Second Model described by equation. 4.11 is experienced by the two fictitious electrons within the hydrogen molecule and varies as a function of the left potential barrier. In this framework, Atom₂ is bound to an atom (Atom₁) located on the solid surface, which is schematically represented by the cyan rectangle. The steepness of the potential barrier is directly related to the electron mobility in the system, which in turn depends on the position of the Fermi level in metals or on the magnitude of the band gap in insulators and semiconductors. Adapted from Ref. [184].

A key element of the model is the introduction of an additional tunable potential barrier. This barrier plays a crucial role in emulating different classes of materials. Within this framework, the effective potential experienced by an electron in the 1D system is given, in atomic units, by:

$$V(x, t) = -\frac{1}{\sqrt{1 + (x - X_1)^2}} - \frac{1}{\sqrt{1 + (x - X_2)^2}} + ax + xE(t) \quad (4.11)$$

where the quantities X_1 and X_2 denote the positions of Atom₁ and Atom₂, respectively, as illustrated in Fig. 4.13. The parameter a controls the strength of the additional potential barrier: by adjusting its value, the system can effectively reproduce behaviors ranging from

metallic to insulating. In particular, when the barrier is absent or negligible ($a \rightarrow 0$), the system behaves like a metal, allowing electrons to move freely across the structure. Conversely, when a steep barrier is present ($a > 0$), electron motion is hindered, mimicking the reduced conductivity characteristic of semiconductors or insulators. In this sense, the barrier acts either as a conductive channel or as an obstacle to electron flow.

It is worth noting that, in a strictly one-dimensional model, the standard Coulomb potential of the form $-1/x$ exhibits a singularity at $x = 0$. This singular behavior leads to strong localization of the electronic wavefunction near the nuclei, significantly increasing the numerical cost and instability of the simulations. To overcome this issue, a softened Coulomb potential of the form $-1/\sqrt{1+x^2}$ has been employed. This modified potential removes the singularity while preserving the correct asymptotic behavior and reproducing the ground-state energy of the hydrogen atom (i.e., -0.5 atomic units). As a result, numerical stability is greatly improved without compromising the essential physics of electron–nucleus attraction.

The process under investigation is the evaporation of Atom₂ from the surface of a solid. In Fig. 4.13), the surface itself is represented by a fictitious aquamarine rectangle, to which Atom₁ belongs, binding Atom₂ at the beginning of each simulation. Both atoms are assumed to be hydrogen-like, each consisting of a single proton and a single electron, thus simplifying the electronic structure while preserving the essential Coulomb interactions. The potential barrier is positioned such that it represents the bulk of the material (the aquamarine rectangle) lying to the left of Atom₁, thereby controlling the degree of electronic coupling between the system and its environment.

As in the case of Si(OH)₄, the evaporation process is triggered by the combined action of a static electric field and a time-dependent THz pulse, as defined in equation 4.3. The static field is always taken to be positive, implying that positively charged particles are driven toward the right-hand side of the simulation box. In this 1D geometry, the spatial coordinate x increases toward the detector, which is conceptually placed at large positive x .

Another important aspect of the model is the linear nature of the added potential barrier. This choice allows for a clear interpretation of the interplay between the barrier and the externally applied static electric field. In particular, if the static field is sufficiently strong, its effect can effectively compensate for the barrier, allowing electrons to overcome it. This situation mimics the behavior of insulating materials under high electric fields, where electrons can be promoted across the band gap into the conduction band, thereby increasing the material's effective conductivity.

Results from the Second Model

The results obtained with this model are presented in Figs. 4.14 and 4.15. In Fig. 4.14, the behavior of the system is compared in the absence and in the presence of the potential barrier. When no barrier is present, a peak electric field of 51.4 V/nm (see Fig. 4.14 (b2)) is sufficient to induce evaporation of Atom₂ under a positive THz pulse, as shown by the light and dark green curves. In contrast, under a negative pulse, Atom₂ does not evaporate but instead oscillates around its equilibrium position, as indicated by the cyan and blue curves. The velocity of the evaporating atom is observed to increase linearly with the strength of the static electric field (light green curve), while remaining approximately constant in the absence of such a field (dark green curve), as shown in Fig. 4.14 (b3).

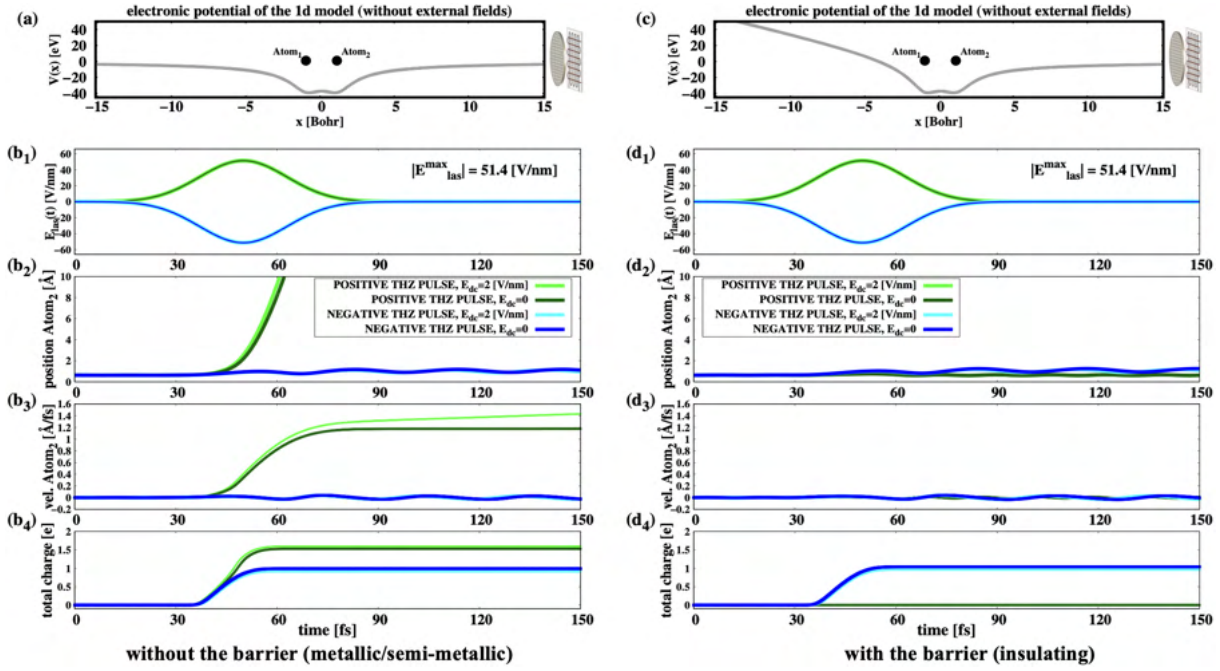


Figure 4.14: Comparison of the results obtained with the two different potentials (metallic - no barrier, insulating - barrier) in (a) and (c) for a fixed amplitude of the THz pulse in (b1) and (d1). The maximum amplitude is 0.1 atomic units (51.4 V/nm the relative field). The green curves correspond to the positive pulse, while the blue lines represent the negative pulse, under presence (light green, cyan) or absence (dark green, dark blue) of an electric field E_{DC} of 2 V/nm. In (b2) and (d2), the position of Atom₂ is shown as a function of time, and in (b3) and (d3), its velocity in the different cases. Finally, (b4) and (d4) show the total charge inside the box. Adapted from Ref. [184].

In the presence of a static field, positive THz pulses induce strong ionization of the system.

This ionization leads to the buildup of positive charge and, consequently, to strong electrostatic repulsion, which drives the evaporation process. Negative pulses, on the other hand, do not produce sufficient ionization and therefore fail to break the bond between Atom₁ and Atom₂. The degree of ionization depends sensitively on the combination of static and time-dependent fields, with the highest ionization occurring for positive THz pulses in conjunction with a static field, as shown in Fig. 4.14 (b4) and as discussed in Section 4.5.1. Up to here, the behavior of this system agrees with what found in the First Model for another molecule, Si(OH)₄.

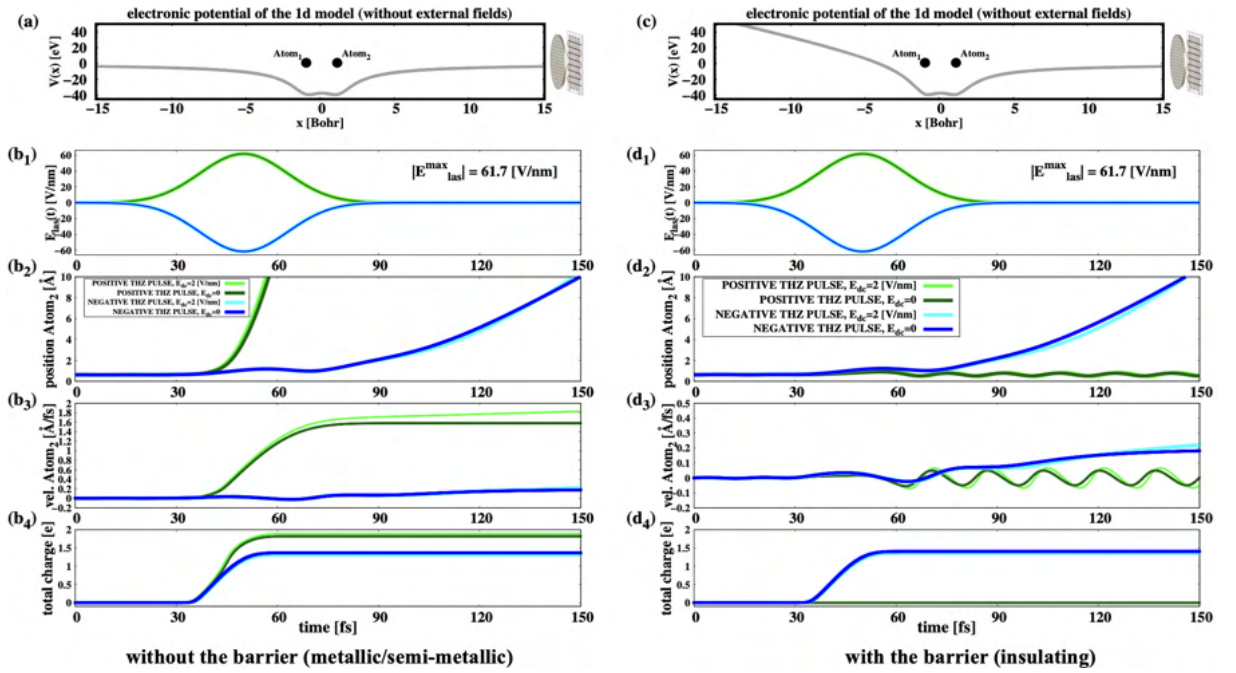


Figure 4.15: Comparison of the results obtained with the two different potentials (metallic - no barrier, insulating - barrier) in (a) and (c) for a fixed amplitude of the THz pulse in (b1) and (d1). The maximum amplitude is 0.2 atomic units (61.7 V/nm the relative field). The green curves correspond to the positive pulse, while the blue lines represent the negative pulse, under presence (light green, cyan) or absence (dark green, dark blue) of an electric field E_{DC} of 2 V/nm. In (b2) and (d2), the position of Atom₂ is shown as a function of time, and in (b3) and (d3), its velocity in the different cases. Finally, (b4) and (d4) show the total charge inside the box. Adapted from Ref. [184].

When the potential barrier is introduced, the system dynamics change significantly. In particular, under a positive THz pulse, evaporation no longer occurs, as shown in Figs. 4.14 (d2) and (d3). This suppression of evaporation is due to the inability of the system to ionize: electrons are prevented from escaping by the barrier and remain localized, resulting in a nearly constant total charge. In contrast, the dynamics under negative pulses are less affected, since

electrons are preferentially driven toward the barrier-free region on the right-hand side, where they can escape more easily towards the detector. As a result, the barrier selectively suppresses electron emission under positive pulses.

This behavior closely resembles that of insulating materials, in which electrons are not free to move across the system. Instead, they respond to external fields primarily through polarization, i.e., through a redistribution of charge without net transport. Consequently, even if the amplitude of the THz pulse is sufficient to induce evaporation in the absence of a barrier, it may become insufficient once the barrier is introduced.

Further insight is provided by the results shown in Fig. 4.15, where a larger amplitude of the time-dependent field is considered. In the absence of the barrier (left panels), Atom₂ evaporates under both positive and negative THz pulses, although the positive remains more efficient. In contrast, when the barrier is present (right panels), evaporation occurs only under negative pulses, while positive pulses fail to induce ionization and subsequent evaporation. This leads to a situation in which the threshold field required for evaporation is lower for negative pulses than for positive ones.

Such asymmetry in the evaporation threshold is in qualitative agreement with experimental observations in silica-based systems, where the polarity of the THz pulse plays a decisive role. The model therefore provides a simple yet physically meaningful explanation of how reduced electron mobility and the presence of potential barriers can give rise to polarity-dependent evaporation dynamics in the case of the THz lasers. In fact, additional simulations reported in Fig. 4.16 demonstrate that the system response to multicycle ultraviolet (UV) fields is not affected by the barrier, as in the case of UV, pulse polarity is not a relevant parameter (see Fig. 4.3).

Also single-cycle infrared (IR) pulses were tested, as can be found in Fig. B.3 in the Appendix B, highlighting the generality of the observed behavior across different frequency regimes.

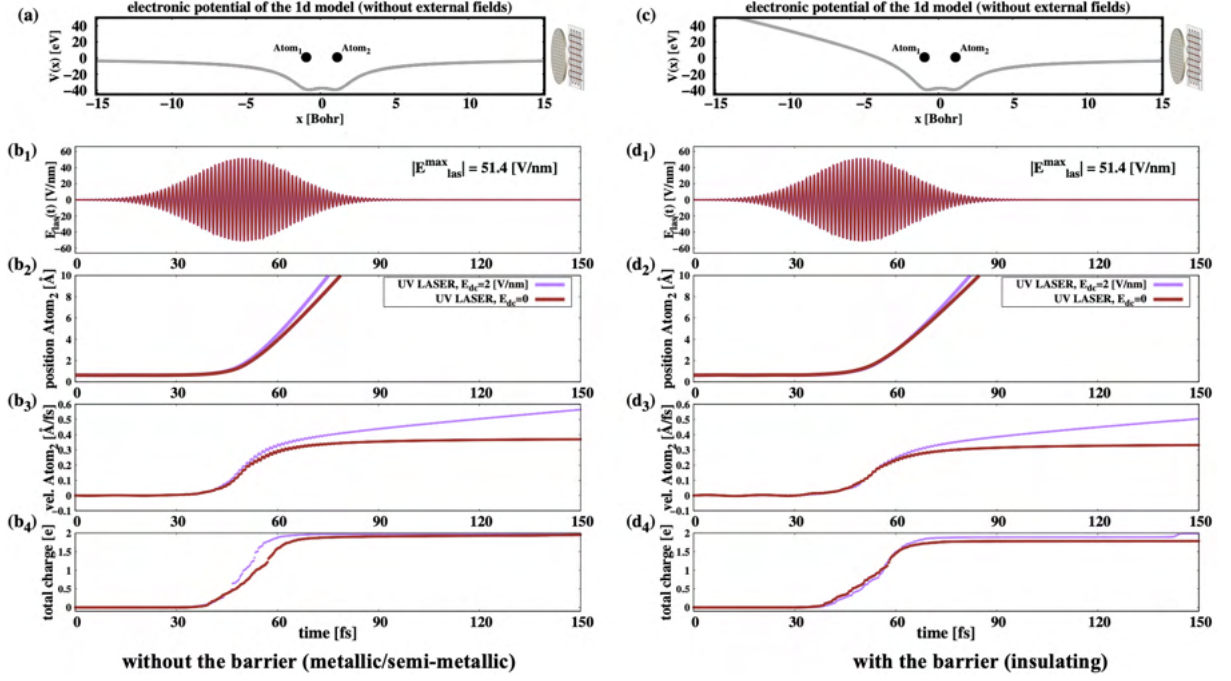


Figure 4.16: Comparison of the results obtained with the two potentials shown in panels (a) and (c) is reported for a fixed amplitude of a UV laser with frequency $\omega = 3.61$ eV in panels (b1) and (d1). The maximum amplitude is 0.1 atomic units, consistent with Fig. 4.14. Panels (b2) and (d2) display the position of Atom₂ as a function of time, while panels (b3) and (d3) show its velocity for the case of a laser combined with a static potential (purple line) and in the absence of an external potential (brown line). Finally, the total charge within the simulation box is reported in panels (b4) and (d4). On the left-hand side (metallic case, panels (a) and (b1–b4)), Atom evaporates when a positive THz pulse is applied (light green and dark green lines), as a consequence of the strong ionization observed in panel (b4). On the right-hand side (insulating case, panels (c) and (d1–d4)), a similar evaporation process is observed even in the presence of a linear potential barrier. In contrast to the THz-driven dynamics, the potential barrier becomes ineffective under UV laser excitation, as electrons can be directly ejected from the surface, as shown on the right-hand side of panels (a) and (c). Adapted from Ref. [184].

4.5.3 The True Role of Electron Dynamics in the Context of Negative Pulse Polarity and APT

Ab initio MD Simulations were essential to get the following results: first of all, THz-assisted APT is always driven by ionization mechanisms. Moreover, in the case of silica, only THz negative polarity pulses can properly induce ionization, since SiO₂, being an insulator, does not allow electrons to flow towards the inside of the silica tip. In other words, the polarity of the THz pulse determines whether the time-dependent field reinforces or opposes the static DC field

at the tip apex. The total electric field is given by the sum of the static and time-dependent contributions, $E_{\text{tot}} = E_{\text{DC}} + E_{\text{THz}}(t)$. In the adopted convention, the static field points toward the detector, i.e. drives positive charges toward the detector. A negative THz pulse corresponds to a field oriented in the opposite direction. When the magnitude of the THz field exceeds that of the static field, i.e. $|E_{\text{THz}}| > E_{\text{DC}}$, the total field reverses direction. As a result, such total field points away from the detector, and electrons, which move opposite to the field due to their negative charge, are accelerated toward the detector and can be detected. Ultimately, this electron loss causes Coulomb repulsion, and positive ions are ejected out of the tip. These ions are then carried by the DC field towards the detector, as they do not sense too much the THz field once left the silica sample.

This means that, for the negative case, both ions and electrons can in principle reach the detector, for two different reasons: the first are driven by Coulomb repulsion and the DC field, the latter by the THz negative pulse. However, in the case of electrons things are a bit more complex. In fact, if the ions typically reach the detector once left the tip, some electrons might come back towards it, as they sense the static field driving them in the opposite direction. This effect is quite relevant here, because these coming-backward electrons represent a key aspect to investigate, ultimately giving rise to the so-called electron-assisted mechanism, that takes place only under negative THz polarity. More details will be provided in the following paragraphs.

As aforementioned, for all cases in which $|E_{\text{THz}}| > E_{\text{DC}}$, electrons are driven towards the detector. Thus, increasing V_{DC} reduces the electrons detected rate until it falls to background, defining a threshold that can be identified as blocking voltage V_{block} . As shown in Fig. 4.17 (a), if $V_{\text{DC}} = 0$ (meaning that $E_{\text{tot}} = E_{\text{THz}}(t)$), electrons leave the tip apex, being subjected only to the negative THz polarity. In this case, a current of electrons flowing out the tip gives rise to a phenomenon called Joule heating. In other words, the kinetic energy of these electrons is influenced by the amplitude of the negative peak of the THz pulse and by the angle at which the electrons are emitted from the tip surface. In a resistive tip, the THz-emitted electrons contribute to apex heating and the power dissipated as heat is proportional to the square of the electron current:

$$P = I^2 R \quad (4.12)$$

or equivalently, per unit volume,

$$q = J^2 \rho \quad (4.13)$$

where I is the electron current, R is the electrical resistance of the heated region, J the current density, and ρ the resistivity of the material. The resulting steady-state temperature rise is, to first order,

$$\Delta T \approx \frac{P}{G_{\text{th}}} = \frac{I^2 R}{G_{\text{th}}} \quad (4.14)$$

where G_{th} is the thermal conductance from the heated region to the base, determined by geometry and material properties. Therefore, a higher electron current produces a larger ΔT , increasing the probability that ions can overcome the energy barrier for evaporation (see equation 1.14, displaying a linear-in- T trend). This means that, THz negative polarity pulse gives rise to small thermal effects, assisting the ion ejection process (from here: electron-assisted mechanism).

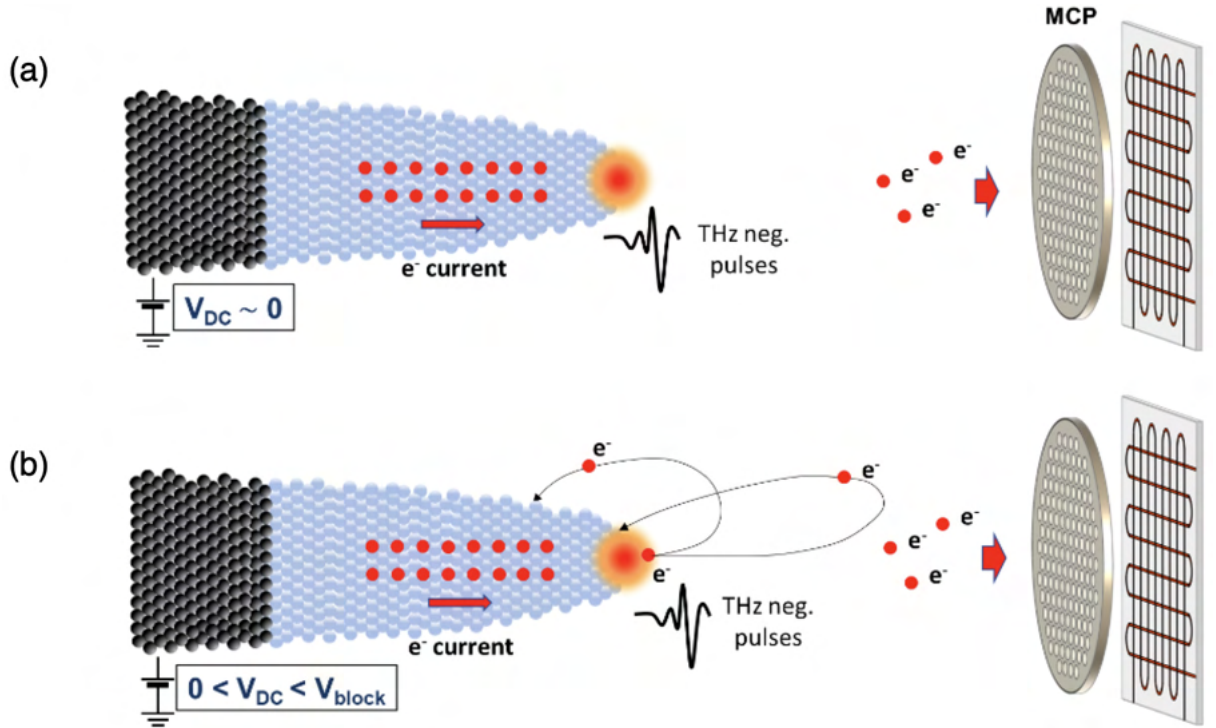


Figure 4.17: Electron-assisted mechanism under negative THz polarity. (a) For $V_{\text{DC}} \approx 0$, THz pulses induce electron emission, and the emitted electrons reach the detector, while the ion rate remains negligible. (b) For $0 < V_{\text{DC}} < V_{\text{block}}$, a fraction of the emitted electrons is decelerated and redirected back toward the tip apex, where they redeposit their energy. This process generates localized heating, which in turn assists ion evaporation. Adapted from Ref. [184]

If instead the DC voltage is such that $0 < V_{\text{DC}} < V_{\text{block}}$, two distinct heating mechanisms contribute simultaneously to the temperature deposition at the tip apex, favoring even more evaporation: *i* the through-tip electron current that dissipates Joule heat and *ii* a significant fraction of the electrons emitted by the THz pulse that does not escape toward the detector, due to the presence of the positive DC bias. In fact, these electrons are decelerated by V_{DC} after emission and can be redirected back toward the tip apex, as illustrated in Fig. 4.17 (b). Upon returning, these electrons impact the surface and deposit their kinetic energy locally. This

process effectively converts part of the electron kinetic energy into localized heating at the apex. The combination of these two mechanisms results in a highly non-uniform energy deposition. In particular, the back-acceleration of electrons concentrates power within a nanometric region near the apex, thereby enhancing local thermal gradients. This localized heating can significantly influence the evaporation dynamics by lowering the effective energy barrier for ion desorption.

The fraction of electrons that return to the tip depends sensitively on both their initial emission angle and the time at which they are emitted within the THz pulse cycle. Since the instantaneous electric field varies rapidly during the pulse, electrons emitted at different times experience different acceleration and deceleration phases, which determines whether they escape or are redirected back to the surface.

Experimentally, this behavior is reflected in the ion detection rate, and was observed by our colleagues from the University of Rouen. The ion rate is often observed to peak in the regime $V_{\text{DC}} < V_{\text{block}}$, where the fraction of returning electrons is maximal and electron-induced heating is most effective. As the DC voltage approaches V_{block} , fewer electrons are able to escape or return to the tip, since emission into free space becomes suppressed. Consequently, the electron-assisted heating mechanism diminishes, leading to a reduction in the ion evaporation rate [184].

A quick sum up of the role that electrons play within the THz-assisted APT (with negative pulse) is displayed here:

- They carry away negative charge, which modifies the local electrostatic environment and can assist in creating conditions favorable for ion emission.
- A fraction of these electrons is decelerated by the DC bias and redirected back toward the apex, depositing kinetic energy in a highly localized region. This generates electron-assisted Joule heating, producing local thermal gradients that enhance the probability of ion evaporation.
- The combined effect of field inversion and electron backscattering explains why negative pulses are far more efficient at promoting cation evaporation in insulating samples than positive pulses.

Notably, from Fig. 4.4 it is clear that THz-negative polarity on sol-gel silica displays some thermal effects, as the spectrum presents thermal tails. As stated in Section 4.1.2, THz lasers should not give rise to significant thermal effects. However, thanks to numerical simulations followed by experimental validations, it was possible to discover that, as in standard UV-assisted APT, the emission under THz drive (specifically under single-cycle negative polarity pulses) also has a benign thermal component as well, that promotes ion evaporation, supporting the main field-driven force of the THz laser.

4.5.4 Near-Apex Field Enhancement in the Context of Negative Pulse Polarity and APT

At this point, a final aspect should be addressed. As stated in Section 4.1.4, in the case of SiO₂ the THz pulse polarity has an effect on the total field sensed by ions and electrons and also on the THz field enhancement near the tip apex. While the first aspect was abundantly covered in the previous sections, the relationship between polarity and field enhancement was not discussed. Intuitively, negative polarity should, given all the presented results, give rise to a better field amplification.

Therefore, to investigate whether positive or negative polarity pulses promote different THz field enhancements, numerical simulations were performed using the commercial software LORENTZ-2E V10.2. Specifically, the THz field enhancement near the apex was measured, given a THz field before amplification of roughly 0.015 V/nm. To recall, the total electric field experienced by electrons is: $V_{\text{tot}}(t) = V_{\text{DC}} + \eta V_{\text{THz}}(t)$. From these calculations, the enhanced THz field at the tip apex was then 20 V/nm for negative pulses and 8 V/nm for positive pulses, implying a field enhancement factor of approximately 1300 for negative and 520 for positive pulses. The same numbers were confirmed by experiments on silica tips (not shown). Taken together, these results demonstrate that negative THz pulses also contribute more significantly to the total field amplification at the tip than positive pulses by a factor of roughly 2.5, corroborating the evidence of their higher efficiency in resolving insulating silica samples. More details about these outcomes can be found in Ref. [184], Fig. 5 (they were not performed by me, but from our colleagues from Rouen).

4.6 Concluding Remarks

This chapter provides a comprehensive overview of a combined theoretical and experimental study aimed at understanding the mechanisms responsible for the evaporation of cations from nanoneedle-shaped tips of materials with differing electrical properties, with particular attention to SiO_2 as a representative insulating material. Experiments were performed using a THz-assisted APT setup, where the mass spectra of insulating SiO_2 samples were compared to those of LaB_6 , a semi-metallic material with high electrical conductivity. In such THz-assisted APT experiments, the total field experienced by the sample consists of a static DC background voltage, V_{DC} , and a single-cycle, time-dependent THz pulse, $V_{\text{THz}}(t)$, oscillating in the THz frequency range.

Single-cycle, high-intensity THz pulses are capable of triggering the evaporation of ions from nonconducting materials, such as amorphous silica, a process that would be inefficient under conventional DC-only conditions due to the poor mobility of electrons in insulators. A critical feature observed in these experiments is the dependence of the evaporation efficiency on the polarity of the THz pulse. Negative pulses induce a significantly stronger response in silica, producing an effective field amplification exceeding a factor of 1300, while positive pulses lead to a reduced amplification, approximately 2.5 times smaller. In contrast, such polarity dependence is absent when multicycle pulses in other frequency ranges, such as UV, are used, indicating that the observed effect is specific to single-cycle high-intensity pulses.

These observations motivated the development of *in silico* approaches to understand the underlying mechanisms. Two models were constructed within an *ab initio* molecular dynamics (MD) framework, employing TDDFT+Ehrenfest dynamics to capture both electronic and nuclear responses under strong static and time-dependent fields.

The First Model focused on a single $\text{Si}(\text{OH})_4$ acid molecule. While it did not fully capture the polarity-dependent effects observed in solids, it highlighted the importance of ionization and electron loss in the evaporation process. Specifically, electron emission during field application reduces the total charge on the system, altering the electrostatic forces that drive cation detachment. However, the limitations of a single-molecule model, particularly in capturing electron mobility and collective effects in extended solids, necessitated a more sophisticated approach.

The Second Model was designed to mimic the interface between a vacuum and a nanoneedle tip. A simplified one-dimensional model with two bound atoms, representing a surface atom and a surface-bound cation, was employed. To simulate insulating behavior, a potential barrier was introduced, which hinders electron motion and suppresses ionization. The simulations revealed that under a positive THz field, evaporation is strongly inhibited in the presence of this barrier, whereas negative pulses efficiently promote electron emission, even when electrons are partially trapped and redirected. This model demonstrates that the polarity dependence of

ion evaporation arises naturally from the interplay between field-induced electron dynamics and the insulating character of the sample.

These two models were essential to capture the following aspects. The asymmetry between negative and positive THz pulses can be attributed primarily to the dynamics of electrons near the tip apex. In the case of a negative pulse, the transient THz field partially cancels the static DC field but can also locally invert the total field, producing a condition where electrons are accelerated toward the detector (as shown by both theoretical models).

Further experimental analysis inspired by computational results showed that electron-assisted heating plays a central role in negative-pulse-assisted ion evaporation. When $0 < V_{\text{DC}} < V_{\text{stop}}$, a fraction of the THz-emitted electrons is decelerated by the static field and redeposited near the tip apex. This backscattering concentrates energy in a nanometric region, creating local thermal gradients that reduce the effective energy barrier for cation emission. Therefore, the combination of field inversion, electron emission, and localized heating under negative pulses provides a coherent explanation for the enhanced ion evaporation observed experimentally in SiO_2 tips.

Finally, the findings from the Second Model indicate that the polarity-dependent behavior observed in THz-assisted APT is not limited to the THz spectral range. Any single-cycle, high-intensity pulse capable of producing sufficient electron emission can, in principle, exhibit similar asymmetry, extending the relevance of these results to other spectral regions.

In conclusion, the combined theoretical and experimental analysis demonstrates that cation evaporation from insulating nanoneedle tips is governed by a subtle interplay of total electric field orientation, electron emission dynamics, and electron-assisted thermal effects. Negative single-cycle THz pulses are particularly effective due to their ability to invert the local field and generate backscattered electrons that locally heat the tip, thus facilitating the evaporation of ions. This study provides a comprehensive framework for understanding polarity-dependent evaporation mechanisms in THz-assisted atom probe experiments.

Understanding The Role of Defects During APT Analyses of Sol-gel Silica through DFT Calculations

This last Chapter contains details about a side project within the MIMOSA, primarily inspired by Ref. [213]. The opening section contains a brief introduction to the link between impurities within silica matrices and optical absorption properties. A second section describes recent experimental advances aimed at studying how impurities change the absorption properties of SiO_2 matrices. These recent outcomes were directly produced by the groups of Angela Vella (University of Rouen - France) and Martin Andersson (Chalmers University of Gothenberg - Sweden), collaborators within the MIMOSA project. These experimental results motivate the use of Density Functional Theory calculations to investigate microscopic optical properties of SiO_2 matrices, providing a simple but effective theoretical baseline. More information about the modeling of optical properties within the DFT pipeline can be found in Chapter 2, Sections 2.6.

5.1 Introduction to the Modeling of the Optical Properties of Defects in SiO_2 Matrices

5.1.1 The Role of Defects within SiO_2 Matrices

Due to its chemical stability, optical transparency, and ability to host a wide range of guest species, silica has been extensively used in applications ranging from glass manufacturing to nanostructured materials. In this thesis silica was mainly described within frameworks concerning biology and Atom Probe Tomography, employing such material for the encapsulation of biomolecular substrates, enabling preservation of the native biomolecule by minimizing mechanical stresses, as shown in previous Chapters. However, SiO_2 matrices present significant challenges for APT [82], that might not concern strictly biological aspects.

In fact, amorphous silica is intrinsically a poor electrical conductor; although its conductivity can increase slightly under specific conditions such as the presence of impurities, intentional doping, elevated temperatures, or exposure to radiation, it remains largely insulating. This

low conductivity complicates APT measurements, as the applied potential transmitted to the specimen apex to generate the high electrostatic field required for field evaporation is typically less enhanced with respect to other materials like metals. Overall, these intrinsic electronic properties pose significant challenges for both experimental characterization and theoretical description.

Silicon dioxide is a wide-band-gap insulating oxide whose electronic structure is dominated by strong covalent bonding between silicon and oxygen atoms. In crystalline polymorphs as well as in amorphous silica networks, the valence band is primarily composed of oxygen $2p$ states hybridized with silicon orbitals, while the conduction band consists mainly of antibonding Si–O states. As a result, the fundamental absorption edge of SiO_2 arises from electronic transitions from the O-derived valence band to the Si–O antibonding conduction band. DFT calculations provide a natural way to investigate these transitions by explicitly computing the electronic band structure and the corresponding transition matrix elements, as described in the theoretical framework reported in Chapter 2, Section 2.6. However, in the case of amorphous or sol-gel derived silica matrices, calculations become more complex because the material does not possess long-range crystalline order. Modelling such systems within DFT therefore typically requires the construction of large periodic supercells that mimic the structural disorder of the amorphous network. These supercells can be generated through MD or Monte Carlo (MC) simulations or structural optimization procedures that reproduce the characteristic bond lengths, bond angles, and coordination environments observed experimentally. Once a representative structural model has been obtained, the electronic structure of the amorphous silica matrix can be computed using standard DFT methods.

An additional level of complexity arises from the presence of structural defects within the silica network. In sol-gel derived materials, defects such as oxygen vacancies, non-bridging oxygen atoms, under-coordinated silicon atoms, or distorted bonding environments may be present as a result of the synthesis process or subsequent thermal treatments. These defects can introduce localized electronic states within the band gap or modify the position of the band edges, thereby altering the optical absorption properties of the material. Within a first-principles framework, the influence of such defects can be investigated by explicitly incorporating them into the atomistic model of the silica matrix and recomputing the electronic structure of the defective system.

The presence of defect states typically leads to new optical transitions that are not present in the pristine material. For example, localized states within the band gap can participate in optical excitations that require lower photon energies than the intrinsic band-to-band transitions of bulk SiO_2 . As a consequence, the calculated imaginary part of the dielectric function may exhibit additional peaks or a shift of the absorption edge toward lower energies. By analysing the

contributions of individual electronic transitions to the dielectric function, it becomes possible to identify which defect states are responsible for specific features in the absorption spectrum. Note that in amorphous silica optical absorption at sub-band-gap energies is dominated by defect states rather than transitions between extended band states. Therefore, accurately capturing the presence and nature of these localized states is more important than obtaining quantitatively precise band edges. Defects can enhance absorption in the UV and deep-UV range by introducing intermediate states that facilitate electronic transitions. This mechanism is particularly relevant for interpreting experimental observations in La-APT, where increased absorption correlates with improved ion yield.

5.1.2 Optical Linear Response and Defects Modeled *via* Density Functional Theory

From a theoretical perspective, understanding the role of defects in amorphous silica requires an accurate description of both structural disorder and electronic localization. Within standard DFT, optical absorption is typically computed using the independent particle approximation, where transitions are calculated between occupied and unoccupied Kohn–Sham states. In other words, DFT enables the identification of defect-induced states, their energetic position within the band gap, and their spatial localization, all of which are essential for interpreting optical absorption mechanisms. As aforementioned, defects can introduce localized electronic states within the band gap of the material. If a defect state lies within the forbidden gap, it can participate in optical transitions involving either the valence band or the conduction band, leading to additional features in the optical absorption spectrum. Specifically, the contribution of defect states to the optical response can be expressed schematically as equation 5.1 (that reminds equations 2.99 and 2.104 in Chapter 2, Section 2.6):

$$\varepsilon_2^{\text{defect}}(\omega) \propto \sum_{i,f} |\langle \psi_f | \hat{\mathbf{p}} | \psi_i \rangle|^2 \delta(E_f - E_i - \hbar\omega), \quad (5.1)$$

where ψ_i and ψ_f denote electronic states associated with defect levels and extended band states. Note that this approach neglects excitonic effects and local field variations, which can be significant in insulating materials (see below).

It is important to emphasize that Eq. 5.1 does not introduce a new formal expression specific to defects, but rather corresponds to the standard independent-particle formulation of the imaginary part of the dielectric function. The effect of defects enters through the modification of the electronic structure, namely by introducing localized states within the band gap or by perturbing the extended states near the band edges. As a consequence, additional optical transitions become possible.

In this context, the defect-related contribution to the optical response can be understood by restricting the sum over transitions to those involving defect states. Schematically, this can be written as

$$\varepsilon_2^{\text{defect}}(\omega) \propto \sum_{\substack{i \in \{\text{VB, defect}\} \\ f \in \{\text{CB, defect}\}}} |\langle \psi_f | \hat{\mathbf{p}} | \psi_i \rangle|^2 \delta(E_f - E_i - \hbar\omega), \quad (5.2)$$

where the sum is restricted to transitions involving at least one defect state. More explicitly, the total imaginary part of the dielectric function can be decomposed into distinct contributions associated with different classes of transitions:

$$\varepsilon_2(\omega) = \varepsilon_2^{\text{VB} \rightarrow \text{CB}}(\omega) + \varepsilon_2^{\text{VB} \rightarrow \text{defect}}(\omega) + \varepsilon_2^{\text{defect} \rightarrow \text{CB}}(\omega) + \varepsilon_2^{\text{defect} \rightarrow \text{defect}}(\omega). \quad (5.3)$$

This decomposition highlights that defects do not alter the formal structure of the optical response, but instead give rise to new absorption channels and redistribute oscillator strength, thereby modifying the spectral features of $\varepsilon_2(\omega)$.

Within this computational framework, the Kohn–Sham orbitals are expanded in a plane-wave basis set and the interaction between valence electrons and ionic cores is described using pseudopotentials. After a self-consistent field calculation determines the ground-state charge density and the corresponding Kohn–Sham eigenstates, the linear optical response of the simulation box containing the silica matrix unit cell can be evaluated in a subsequent post-processing step. This step involves calculating momentum or velocity matrix elements between valence and conduction states and performing the Brillouin zone integration required to obtain the dielectric tensor. In practice, the Dirac delta function appearing in equation 5.1 is replaced by a finite broadening function, which accounts for the finite lifetime of electronic excitations and produces a continuous absorption spectrum comparable to experimental measurements. See Chapter 2, Section 2.6 for more details, also regarding how defect states are typically characterized within DFT and linked to absorption spectra. In summary, the most common observables that can be extracted from simulations to reconstruct the absorption spectra are the **Density of States (DOS)**, identifying the presence of states within the band gap, and the **Projected Density of States (PDOS)**, determining the atomic and orbital character of defect states. Localized states associated with defects often appear as sharp peaks within the band gap and are spatially confined around specific atomic sites. These states play a dominant role in sub-band-gap optical absorption.

The predictive capability of DFT is limited by its ground-state nature. Standard exchange–correlation functionals tend to underestimate the band gap of silica, which in turn affects the predicted position of defect states and the description of optical transitions. Despite these limitations, DFT remains a powerful and widely used approach for studying defect physics in

amorphous materials, particularly when the focus is on qualitative trends and the identification of localized states rather than quantitatively exact excitation energies. The accuracy of DFT calculations depends critically on the choice of the exchange–correlation functional. Common approximations include the Local Density Approximation (LDA) and the Generalized Gradient Approximation (GGA), such as the Perdew–Burke–Ernzerhof (PBE) functional [158]. While these functionals provide a reasonable description of ground-state properties, they systematically underestimate band gaps due to the lack of proper quasiparticle corrections. The underestimation of the band gap has important consequences for defect studies. Defect levels that should lie deep within the band gap may appear artificially close to the conduction band, affecting both their energetic position and their optical activity, specifically in wide-band-gap materials such as silica, where discrepancies can exceed several electronvolts [214].

Moreover, standard DFT does not explicitly account for electron–hole interactions, which are essential for accurately describing optical absorption processes, as described in Section 2.6.3. As a result, calculated absorption spectra may lack quantitative accuracy, especially near defect-induced transitions. More advanced methods, such as the GW approximation and the Bethe–Salpeter Equation (BSE), provide improved descriptions of electronic excitations and optical spectra. However, their computational cost scales poorly with system size, making them impractical for large amorphous models that need supercells. In addition, the presence of disorder and a high density of localized states complicates the convergence of these methods. Luckily, optical absorption in the context of amorphous SiO₂ is often dominated by defect-related states within the band gap rather than by the precise energies of extended band states. Thus, capturing the presence, nature, and spatial localization of these defect states is more critical than achieving quantitatively exact excitation energies. Practically, this approximation assumes that microscopic variations of the induced electric field within the unit cell can be ignored, so that the electrons respond only to the macroscopic average field. For these reasons, standard DFT remains the method of choice for large-scale simulations of defective amorphous silica.

Overall, the combination of DFT and linear response theory therefore provides a powerful first–principles framework for modeling the optical properties of complex oxide materials. By computing the electronic eigenstates of the system and evaluating the transition probabilities between occupied and unoccupied states, it becomes possible to predict the frequency–dependent dielectric function and the corresponding absorption spectrum directly from the atomic structure of the material. When applied to realistic structural models of silica matrices, including those containing defects and structural disorder, this approach enables the identification of the microscopic mechanisms responsible for experimentally observed spectroscopic features, thereby establishing a direct connection between atomic structure, electronic properties, and optical response in SiO₂ materials. In other words, a direct connection between computational calcula-

tions and experimental spectroscopy can be established and exploited.

5.2 Motivating DFT Calculations: Experimental Data on Defects of SiO₂ Matrices within the APT Context

Recently, the groups of Angela Vella (University of Rouen - France) and Martin Andersson (Chalmers University, Gothenburg - Sweden) collaborated with us to unravel the role of defects within SiO₂ during APT measurements [213]. In this section, the experimental results that motivated the employment of DFT calculations will be explained.

Specifically, the work done by our collaborators within the MIMOSA project focuses on the influence of laser wavelength on the APT analysis of silica, correlating evaporation conditions, mass spectral features, thermal effects, and compositional defects of the matrix. A systematic comparison between deep-UV (257.5 nm, LEAP 6000 XR), UV (343 nm, La-WATAP – Cameca), and green (515–532 nm, La-WATAP - Cameca) laser illumination reveals the fundamental role of optical absorption and electronic excitation in controlling the field evaporation process in this highly insulating material.

Notably, we are interested in only a partial part of the results, concerning the defects in the APT silica matrices and not the comparison UV/deep-UV lasers in terms of evaporation efficiency. Thus, the following sections will be mainly focused on the former aspect, whereas the latter will be instead poorly treated (more details about it can be found though in Ref. [213]).

5.2.1 SiO₂ Matrices in APT Display Defects

A first important observation concerns the feasibility of APT measurements under these conditions. Analyses performed using green light were generally unsuccessful, as the samples systematically fractured before a sufficient number of ions (typically $< 10^5$) could be collected. This behavior is directly related to the low optical absorption of silica at these wavelengths. Because the material absorbs poorly in the green spectral range, the incident laser energy is not efficiently converted into localized electronic excitation. Instead, it leads to excessive and spatially delocalized heating, which promotes mechanical failure before a stable evaporation regime can be established.

In contrast, the use of UV and deep-UV illumination significantly improves the success rate of the analyses, as displayed in Figs. 5.1 (a) and (b). Under these conditions, a lower onset voltage is sufficient to reach the target evaporation rate, allowing the acquisition of datasets containing millions of ions. The improved performance with respect to green light can be attributed to the higher absorption coefficient of silica at shorter wavelengths, which enables a more efficient coupling between the laser pulse and the electronic system of the material.

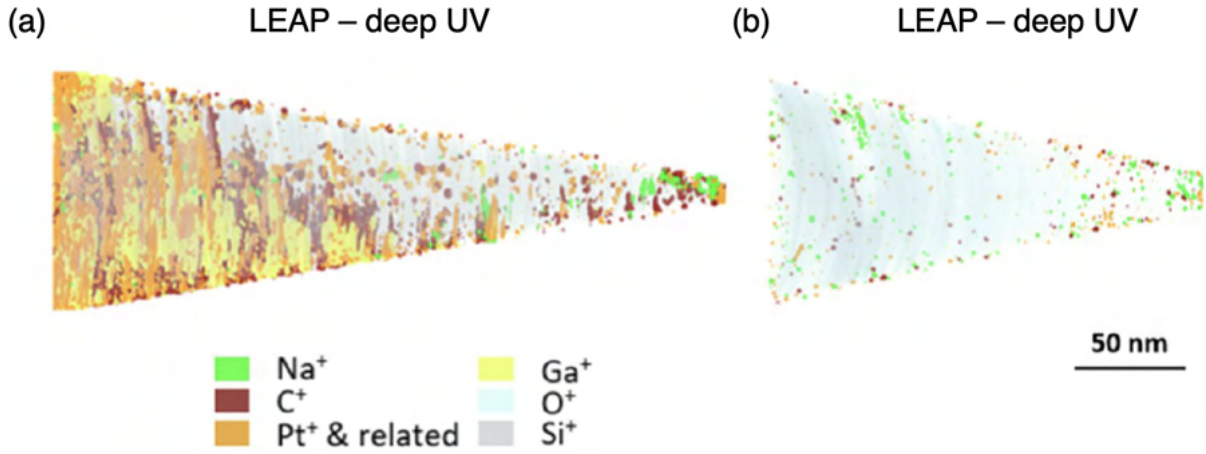


Figure 5.1: Three-dimensional atomic reconstruction of the analyzed silica samples, showing the spatial distribution of ionic species. Silicon and oxygen are represented as gray and cyan dots, respectively, while carbon, sodium, gallium, and platinum are shown as iso-surfaces in dark brown, green, yellow, and orange, with composition thresholds of 18%, 12%, 45%, and 25%, respectively. These latter species originate from the FIB-SEM sample preparation procedure. The reconstructions were obtained from (a) deep-UV (LEAP) laser-assisted atom probe tomography (La-APT), using a starting radius of 15 nm and a cone angle of 6.5°, and (b) UV (LaWATAP) La-APT, using a starting radius of 15 nm and a cone angle of 15°. Adapted from Ref. [213].

Sodium (green in Fig. 5.1), deriving from the matrix precursor (sodium silicate), is also detected throughout the reconstructed volume. Its concentration is slightly higher at the tip apex, suggesting a redistribution process. This behavior is consistent with field-driven migration of alkali ions. The presence of clusters of Pt and C within the silica matrix is also captured. These species are introduced during sample preparation FIB analysis and are found to infiltrate the porous structure of the material. A notable difference between deep-UV and UV analyses concerns diffusion. Under deep-UV illumination, the distribution of Pt, Ga, and C appears more diffuse, indicating enhanced atomic mobility. This might be attributed to the stronger laser-induced heating, which facilitates diffusion processes. Additionally, when the silica region is small, direct irradiation of the Pt weld can further increase local temperatures, assisting evaporation (lowering the relative evaporation barrier of ions).

The mass spectra obtained under deep-UV and UV illumination exhibit similar dominant features. The most intense peaks are observed at 14 Da and 16 Da, corresponding to Si²⁺ and O⁺ ions, respectively. Additional peaks are identified at 28 Da (Si⁺), 32 Da (O₂⁺), 44 Da (SiO⁺), and 60 Da (SiO₂⁺). Some smaller contributions at 22-23 Da correspond to sodium originating from the precursor solution. Intriguingly, there is also a consistent detection of Ga, Pt, C species (also present in the resolved 3D tip in Fig. 5.1). These contaminant atoms constitute structural

defects that should be further analyzed, as they might modify the SiO_2 properties. To recall, the more conductive the silica, the more the field enhancement near the tip apex, the more effective the APT analysis.

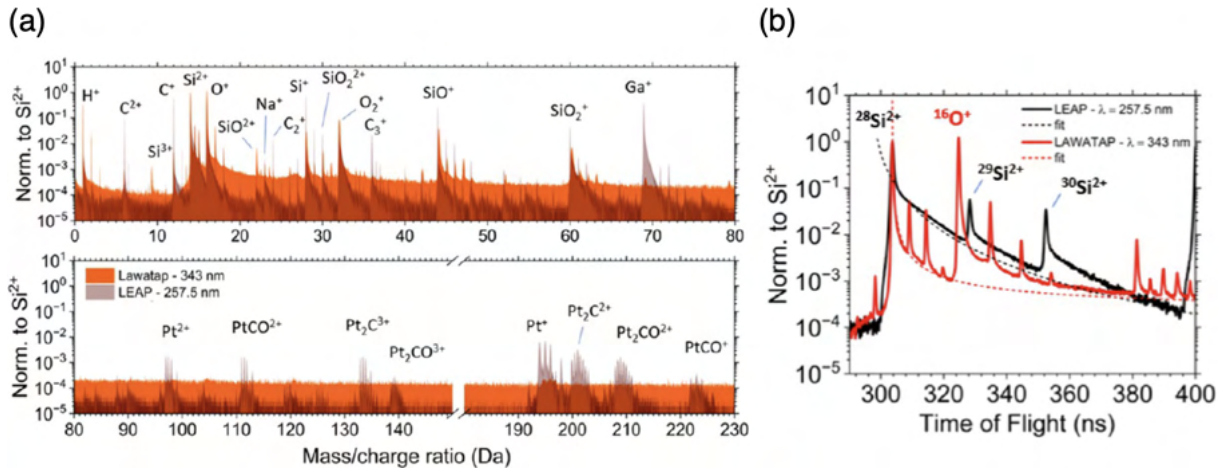


Figure 5.2: (a) Mass spectra of silica samples measured in the ranges [0–80] Da (upper) and [80–230] Da (lower) using deep-UV (dusty pink spectrum) and UV (orange spectrum) illumination. For deep-UV, 17 million ions were collected at a laser energy of 250 pJ (energy density 80 J/m^2) with a detection rate of 0.5% ions/pulse at a base temperature of $T = 50 \text{ K}$. For UV, 17 million ions were collected at 50 nJ (energy density 70 J/m^2) with a detection rate of 0.25% ions/pulse at $T = 50 \text{ K}$. (b) time-of-flight (TOF) spectrum from the APT analyses of silica samples using deep-UV (black line) and UV (orange line) illumination. The TOF spectra in were fitted using the cooling equation reported in Ref. [213], yielding cooling times of 7.35 ns for deep-UV and 0.5 ns for UV analyses. Notably, deep-UV conditions promote a slower thermal relaxation of the tip, primarily due to the smaller cone angle and resulting reduced heat dissipation, which leads to longer cooling times and broader peak tails with respect to UV. Adapted from Ref. [213].

5.2.2 Defects Alter the Absorption Properties of the SiO_2 Matrices utilized in APT

A second analysis concerns the measuring of absorption properties of silica matrices utilized in APT through UV/Vis spectroscopy. The absorbance was measured with a Hewlett- Packard 8453 UV/Vis spectrometer, which was blanked against air. The absorption spectrum obtained is shown in Fig. 5.3 and exhibits a nearly constant absorbance above 400 nm, which can be partially attributed to light scattering due to the non-ideal flatness of the sample. However, more relevant features appear at shorter wavelengths, where two distinct increases in absorbance are observed around 400 nm and, more prominently, around 300 nm.

A key quantitative result is that the absorbance increases by approximately a factor of three when moving from 343 nm (UV) to 258 nm (deep-UV). Since these photon energies lie below the fundamental band gap of silica, this behavior cannot be explained by intrinsic band-to-band transitions. Instead, it provides direct experimental evidence for sub-bandgap absorption mechanisms, which are typically associated with defect-induced electronic states.

This experimental observation directly correlates with the APT results. Despite similar incident laser energy densities (70–80 J/m²), deep-UV illumination enables stable evaporation at significantly lower electric fields compared to UV illumination (not shown here). This implies that the energy deposited into the electronic system is more efficiently absorbed in the deep-UV range. Consequently, more effective electronic excitation leads to enhanced energy transfer to the lattice, increasing the local temperature at the tip apex and assisting field evaporation. In other words, a higher absorption leads to the generation of more free charges and thus to a higher energy transfer into the structure, which in turn leads to a higher temperature of the silica matrix. This also means that an extremely high electric field is not required to evaporate the surface atoms, reducing the risk of sample breakage. When many free charges diffuse in the sample, the resistivity of the silica also decreases, so the potential drop along the tip is also reduced due to the resistivity effect. This increases the effective electric field at the surface and thus evaporate the atoms more easily.

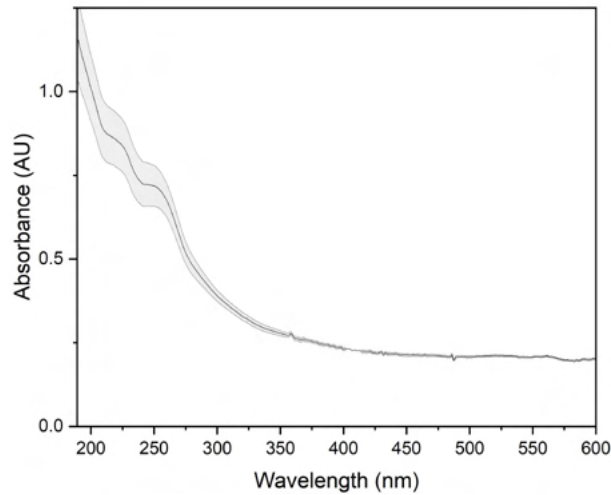


Figure 5.3: Absorption spectrum of a 0.6 mm thick silica sample that has been blanked against air. The standard deviation is shown in grey. Adapted from Ref. [213].

5.3 The DFT Study

The combination of experimental measurements and first-principles simulations a key aspect within this thesis. As stated in the previous sections, first-principles DFT calculations combined with linear optical response theory provide a complementary theoretical perspective by allowing the identification of defect-related electronic states and their corresponding spectroscopic signatures. While DFT has well-known limitations in describing excitation energies, its ability to capture the main structural and electronic characteristics of defect states makes it particularly valuable for studying sub-band-gap optical absorption effects.

In this context, the present work aims to employ DFT-based simulations to investigate the structural, electronic, and optical properties of defective amorphous silica, by comparing results from both pristine and defect-containing matrix models. This approach allows to identify the spectroscopic fingerprints associated with particular defect types and to assess how variations in defect concentration or local structural disorder influence the optical properties of the material. In this way, it should be possible to link the presence of defects in the SiO_2 samples to the enhanced optical absorption observed experimentally, which is strongly connected to APT performances, as explained in the last section.

Typically, in APT matrices lots of defects can be found. Point defects involving oxygen and hydrogen are among the most common in SiO_2 matrices. Oxygen vacancies can be generated by the removal of bridging oxygen atoms, leading to the formation of undercoordinated silicon centers, while hydrogen is typically incorporated as an interstitial species or bonded to oxygen, forming silanol (Si-OH) groups. These defects are particularly prevalent near surfaces and in sol-gel-derived silica, where incomplete network formation and exposure to ambient conditions promote their formation.

The needle shaped specimen is typically mounted to a Tungsten pre-tip with a glue that can contain Carbon or Gallium impurities. These atoms might shift from the support to the silica matrix, originating there several defects within the matrix network, that can be substitutional or interstitial (see below). In particular, the use of FIB for lift-out can lead to substitution or displacement of atomic constituents by the implantation of Ga^+ ions, which can lose all their kinetic energy through elastic and inelastic interactions in the sample and be adsorbed; depending on the Ga ion concentration, this can lead to significant changes in the structural and optical properties of the sample. Lastly, residual Na atoms can be incorporated as interstitial species, often deriving from the initial silicate sodium solution utilized to initiate the sol-gel polymerization process.

For all these reasons, several types of defects were considered, including hydrogen and oxygen defects, carbon impurities, residual sodium, or also gallium at different concentrations. Specifically, this work focuses on two classes of point defects:

- **substitutional defects**, in which a host atom of the silica matrix –either silicon or oxygen– is replaced by a foreign atom, such as carbon or gallium (e.g. $\text{Si} \rightarrow \text{Ga}$), while preserving the original lattice site.
- **interstitial defects**, where an atom of the same or a different chemical species (e.g. Sodium) occupies an interstitial position within the silica network, i.e. a site that is normally unoccupied in the ideal crystal structure. For example, interstitial gallium defects correspond to gallium atoms residing at non-lattice sites within the silica framework.

Different SiO_2 matrices were also considered in this work. In fact, the optical absorption spectra for three diverse DFT-optimized models of pure amorphous silica (see Fig. 5.4) were evaluated in combination of the defect conditions aforementioned. All silica structures are characterized by a cubic box of 1.426 nm size with 192 atoms at a density of 66.2 atoms/nm³, representing relative minima in the potential energy landscape. Also other box sizes were investigated, but only results from systems displaying a 1.426 nm cubic box will be reported.

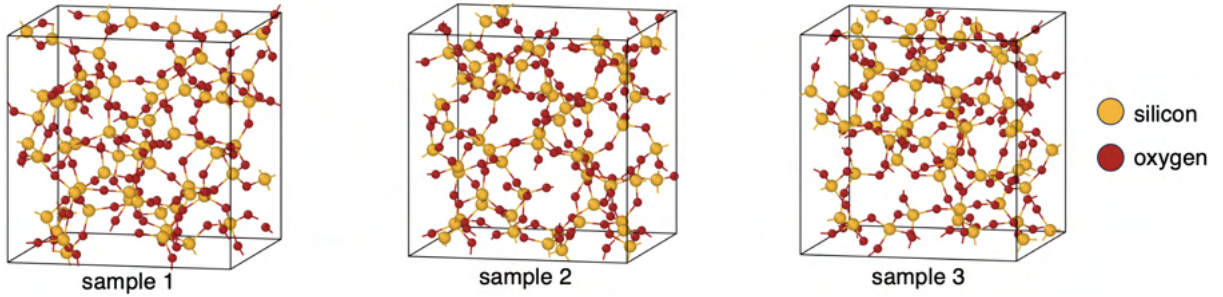


Figure 5.4: Atomistic models of defect-free SiO_2 simulation supercells containing $N = 192$ atoms, corresponding to a cubic box with edge length $L = 1.42$ nm. Oxygen atoms are shown in red and silicon atoms in yellow. Adapted from Ref. [213].

5.4 Simulation Protocol

This section describes the simulation protocol utilized for every DFT calculation, within the Quantum ESPRESSO package [156].

5.4.1 DFT and Optical Properties

The optical properties of the analyzed systems are evaluated within the framework of linear response theory by calculating the imaginary part of the complex dielectric tensor, $\varepsilon_{\alpha\beta}^{(2)}(\omega)$, where α and β denote the tensor components with respect to a set of Cartesian axes. Its spatial

average is directly related to the absorption coefficient, up to a proportionality factor. In the present calculations, local field effects (LFE) are neglected, and the expression for $\varepsilon_{(2)}^{\alpha\beta}(\omega)$ is given by equation [215, 216], similar to equation 2.99 but discretized:

$$\varepsilon_{\alpha\beta}^{(2)}(\omega) = \frac{4\pi^2 e^2}{\Omega N_k m^2 \omega^2} \sum_{n,n',\mathbf{k}} M_{\alpha\beta}^{nn'}(\mathbf{k}) \delta(E_{n'\mathbf{k}} - E_{n\mathbf{k}} - \hbar\omega), \quad (5.4)$$

where e is the electron charge, Ω is the volume of the simulation cell, N_k is the number of $\mathbf{k}$ -points used to sample the Brillouin zone, and m is the electron mass. The quantity $M_{\alpha\beta}^{nn'}(\mathbf{k})$ represents the squared matrix element connecting the initial and final electronic states via the perturbation induced by the electromagnetic field. The transitions occur between states belonging to different bands n and n' at the same crystal momentum $\mathbf{k}$, with energies $E_{n\mathbf{k}}$ and $E_{n'\mathbf{k}}$, respectively. This expression is derived in the long-wavelength limit. The Dirac delta functions are broadened using a small smearing parameter in order to avoid unphysical transitions with infinite lifetimes. See Chapter 2, Section 2.6 for more details.

5.4.2 Exchange–Correlation Functional Utilized

The calculation of the dielectric tensor $\varepsilon_{(2)}^{\alpha\beta}(\omega)$ was carried out as follows. The ground-state electronic density was obtained using a plane-wave basis set with an energy cutoff of 80 Ry and the Perdew–Burke–Ernzerhof (PBE) exchange–correlation functional [158]. Within the same functional, the effect of an external constant electric field on the electronic properties of pristine amorphous silica was also investigated, revealing negligible modifications for all configurations considered.

To improve the description of the electronic band gap, hybrid functional calculations were subsequently performed using the Heyd–Scuseria–Ernzerhof (HSE06) functional [217]. This approach yields a band gap of approximately 8 eV for amorphous silica (see Fig. 5.5), in good agreement with experimental values reported in the range of 8–9 eV [218, 219, 220].

The Heyd–Scuseria–Ernzerhof (HSE) hybrid functional is based on a range separation of the Coulomb interaction into short-range (SR) and long-range (LR) components using the error function:

$$\frac{1}{r} = \frac{\text{erfc}(\omega r)}{r} + \frac{\text{erf}(\omega r)}{r}, \quad (5.5)$$

where ω is the range-separation parameter. In the HSE functional, ω typically defines the range separation of the Coulomb interaction into short-range and long-range components. erfc and erf denote the complementary error function and the error function, respectively. The first term represents the short-range contribution, while the second term corresponds to the long-range part. The parameter ω controls the spatial extent of the short-range interaction. Physically, it

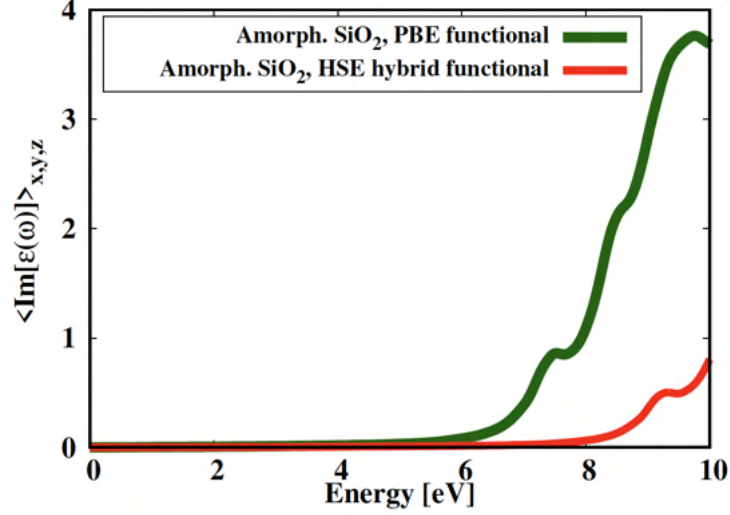


Figure 5.5: Effect of the DFT functional: average of the imaginary part of the dielectric tensor along the orthogonal Cartesian directions. Amorphous silica matrix using a PBE exchange-correlation functional (dark green colour) compared to a hybrid HSE06 functional (red colour).. Adapted from Ref. [213].

defines a screening length $r_{\text{screen}} \sim 1/\omega$, beyond which the Coulomb interaction is treated at the semilocal level. In the standard HSE06 parametrization, $\omega = 0.11 \text{ bohr}^{-1}$, corresponding to a screening length of approximately 5 Å. As a result, exact (Hartree–Fock) exchange is included only at short distances, where electronic interactions are strongest and localization effects are most important, while long-range interactions are described using the PBE exchange functional. This screened hybrid approach significantly reduces the computational cost compared to full-range hybrid functionals, while retaining an improved description of localized electronic states, such as those associated with defects.

Within this framework, the exchange–correlation energy is written as:

$$E_{xc}^{\text{HSE}} = \alpha E_x^{\text{HF,SR}}(\omega) + (1 - \alpha) E_x^{\text{PBE,SR}}(\omega) + E_x^{\text{PBE,LR}}(\omega) + E_c^{\text{PBE}}, \quad (5.6)$$

where:

- α is the fraction of exact exchange,
- $E_x^{\text{HF,SR}}$ is the short-range Hartree–Fock exchange,
- $E_x^{\text{PBE,SR}}$ and $E_x^{\text{PBE,LR}}$ are the short- and long-range components of the PBE exchange,
- E_c^{PBE} is the PBE correlation functional.

In the standard HSE06 parametrization, the parameters are $\alpha = 0.25$, $\omega = 0.11 \text{ bohr}^{-1}$. The short-range Hartree–Fock exchange term is explicitly given by:

$$E_x^{\text{HF,SR}} = -\frac{1}{2} \sum_{ij}^{\text{occ}} \iint \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}) \frac{\text{erfc}(\omega|\mathbf{r} - \mathbf{r}'|)}{|\mathbf{r} - \mathbf{r}'|} \psi_j^*(\mathbf{r}') \psi_i(\mathbf{r}') d\mathbf{r} d\mathbf{r}'. \quad (5.7)$$

The use of the screened Coulomb interaction in the exchange term ensures that exact exchange is included only at short range, which significantly reduces the computational cost while preserving an improved description of localized electronic states.

5.4.3 External Electric Field

A constant electric field of 20 V/nm was applied along the horizontal direction during DFT calculations. This field strength is comparable to experimentally estimated values derived from the charge-state ratio (CSR) in APT analyses performed by our colleagues within the MIMOSA Project [213].

5.4.4 Generation and Optimization of Amorphous Structures and Defective Structures

In atomistic simulations of disordered materials, it is standard practice to initialize the system from a well-defined, defect-free crystalline structure in order to avoid introducing uncontrolled artifacts at the outset. In the present work, β -cristobalite is selected as the starting configuration. This phase is a high-temperature crystalline polymorph of SiO_2 , composed of a network of corner-sharing SiO_4 tetrahedra. Its local tetrahedral coordination closely resembles the short-range order of amorphous silica, making it a suitable precursor for generating realistic glassy structures.

The atomic number density is fixed to $0.0662 \text{ atoms}/\text{\AA}^3$, corresponding to a mass density of approximately 2.2 g/cm^3 , in agreement with experimentally measured values for amorphous silica. This choice ensures that the resulting structures are physically meaningful and representative of amorphous materials. Starting from an ideal periodic β -cristobalite configuration at this density guarantees the absence of pre-existing defects and provides a controlled initial condition for subsequent amorphization.

The crystalline structure is then transformed into an amorphous network through a melt–quench and relaxation procedure. In this work, the Monte Carlo (MC) approach introduced by Wooten, Winer, and Weaire is employed to efficiently generate defect-free amorphous configurations [221]. In this method, the atomic structure is represented as a continuous random network, and the bonding topology is modified through a sequence of bond-switching moves. Initially, a series of

random bond rearrangements is performed to disrupt the crystalline order, effectively randomizing the network.

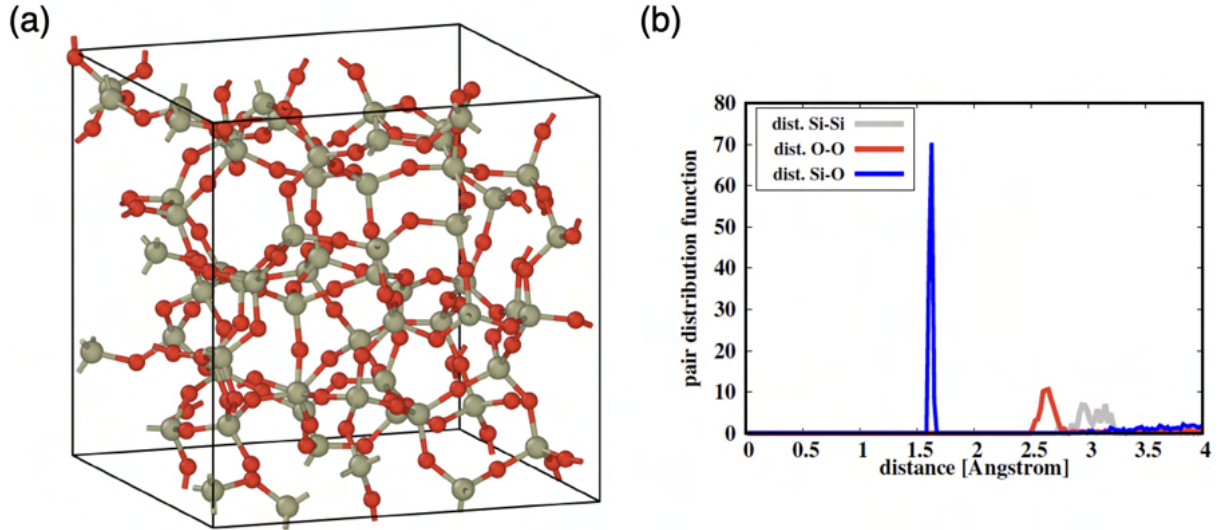


Figure 5.6: Model of the amorphous silica matrix: (a) SiO_2 simulation cell containing $N = 192$ atoms, corresponding to a cubic box with edge length $L = 1.42$ nm. The structure is defect-free; oxygen atoms are shown in red and silicon atoms in grey. (b) Corresponding pair distribution function. Adapted from Ref. [213].

Following this randomization step, the system is annealed at constant volume by progressively lowering the temperature. At each step, atomic displacements and bond rearrangements are proposed and accepted or rejected according to a Boltzmann probability criterion, based on the associated change in potential energy. The energy evaluation is carried out using an empirical force field parametrized for SiO_2 [222], ensuring a realistic description of interatomic interactions. This approach is conceptually similar to molecular dynamics-based annealing procedures, but offers improved efficiency in exploring configurational space [223, 224].

The Wooten–Winer–Weaire (WWW) algorithm enables the generation of fully amorphous structures while preserving the fourfold coordination of silicon atoms, thereby avoiding the introduction of coordination defects. As a result, the obtained configurations correspond to low-energy, defect-free amorphous silica networks that are well suited for subsequent electronic structure calculations.

The final amorphous structures are contained within cubic simulation cells with periodic boundary conditions. To assess finite-size effects, three different system sizes are considered, corresponding to cubic boxes with edge lengths of $L = 10.283$, 14.26 , and 21.39 Å. These models provide a consistent basis for investigating the structural and electronic properties of amorphous

SiO₂ across different length scales. Since results did not vary too much, only the 14.26 box will be presented in the results, represented in Fig. 5.6.

At this point, defects can be introduced into amorphous silica models in several ways:

- **Oxygen vacancies:** Created by removing an oxygen atom, leading to the formation of Si-Si bonds (E' centers). This will be named as **Oxygen defect**.
- **Dangling bonds:** Generated by breaking Si-O bonds without recombination; in some cases, an Hydrogen atom can be inserted close to the Oxygen atom, so that a silanol group forms (OH), leaving the Si atom without a bond. This last case takes the name of **Hydrogen defect**.
- **Interstitial:** Additional atoms inserted into the network, like Ga or Na.
- **Substitutional:** Replacing a Silicon atom with a C or Ga atom.

After introducing defects, optimization procedures like full structural relaxation is performed to allow the system to reach a local energy minimum. Notably, for defective silica structures, only the atomic positions were optimized, while the simulation supercell was kept fixed in order to preserve the target density. As a result, the cubic box is characterized by fixed lattice vectors throughout the relaxation procedure. The structural optimization was performed using a Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm [156], with convergence thresholds of 10^{-5} Hartree/Bohr for the forces and 10^{-6} Hartree for the total energy.

5.5 Results of Optical Absorption Evaluations through DFT

This section focuses on the results deriving from DFT calculations performed on several SiO_2 amorphous matrices containing different defect types. Details about the creation of silica model structures can be found in Section 5.4.4. To recall, pure silica has a very large band gap, but in the context of APT, three effects might reduce it: first, the presence of defects might give origin to new energy levels inside the band gap, populated by electrons receiving energy from the external field. Second, the presence of such strong field lowers the energy difference between occupied and unoccupied states (a weak band-gap narrowing). Third, phonon excitation and/or lattice distortion might arise due to the heating of the sample during irradiation. The bandgap energy of insulators such as silica decreases with increasing temperature, as the amplitude of atomic vibrations increases, leading to larger interatomic distances and thus weaker orbital interaction. Note that, when many free charges diffuse in the sample (i. e. smaller bandgap), both the temperature and the effective electric field at the tip surface are increased and ions evaporate more easily. By systematically introducing representative defects into realistic amorphous models, the goal of this Chapter is to understand the role that defects and impurities play in modifying the optical absorption properties of silica, specifically focusing on the ultraviolet (UV) and deep-ultraviolet (DUV) regions, spanning photon energies from 3.1 to 6.2 eV, and from 6.2 to 12.4 eV, respectively.

5.5.1 Preliminary Analyses on the Amorphous SiO_2 Models

Convergence Analysis on the Box Size of the amorphous SiO_2 Models

When analyzing with DFT amorphous silica bulks, some aspect must be carefully analyzed: first, the absence of long-range order necessitates the use of large supercell models to capture the inherent structural disorder. Typically, supercells containing hundreds of atoms are often required to capture realistic disorder and defect distributions. In our case, three types of periodic, cubic supercells were created, with $L = 10.28, 14.26$ and 21.42 Å, containing amorphous SiO_2 . Second, the diversity of possible defects—ranging from intrinsic point defects to extrinsic impurities such as carbon, hydrogen, or metal ions—requires systematic investigation of multiple configurations. For this reason, the following results have three replicates, as shown in Fig. 5.4.

To obtain a convincing interpretation of the role of defects in the absorption properties of amorphous silica in the APT setup, the DFT-derived optical absorption spectra of the three different optimized supercells were calculated, in order to perform a convergence analysis with respect to the cell size: the main objective is to find the minimum cell box edge, in order to retain physical accuracy while lowering computational costs.

Fig. 5.7 displays the imaginary part of the dielectric tensor, $\varepsilon_2(\omega)$ for the three amorphous

SiO₂ models of edge 1.03 nm ($N = 72$ atoms - red), $L = 1.42$ nm ($N = 192$ atoms - green), and $L = 2.14$ nm ($N = 648$ atoms, yellow). The spectra exhibit a clear dependence on the system size, particularly for the smallest cell: a spurious peak appears in the energy range between 7 and 8 eV, which is absent in the larger simulation cells. This feature could be attributed to finite-size artifacts, artificial interactions between periodic images insufficient sampling of the disordered network, which can lead to unphysical electronic transitions. In contrast, the spectra corresponding to the larger supercells ($L = 1.42$ nm and $L = 2.14$ nm) are qualitatively consistent. Although slight differences are observed in the fine structure of the main absorption peak around 10 eV, these variations can be ascribed to the intrinsic structural disorder of the amorphous network rather than finite-size effects. Importantly, the fundamental band gap remains unchanged between these two models, indicating that the electronic structure is well converged with respect to system size.

These results demonstrate that a supercell with $L = 1.42$ nm ($N = 192$ atoms at a density of 66.2 atoms/nm³) is sufficient to accurately describe the optical properties of amorphous silica within the present computational framework. Therefore, the $N = 192$ model represents an optimal compromise between accuracy and efficiency for the calculation of $\epsilon_2(\omega)$ in amorphous SiO₂.

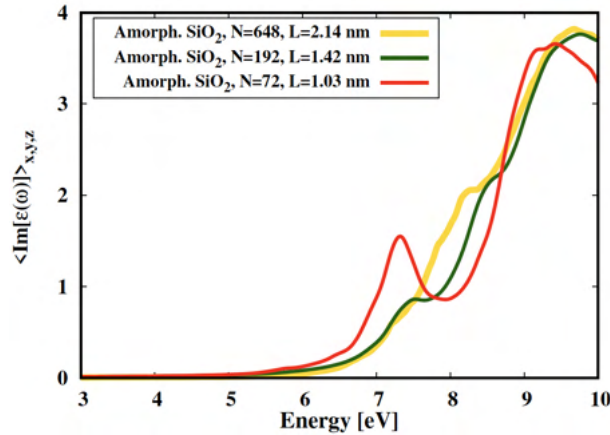


Figure 5.7: Imaginary part of the dielectric tensor for three amorphous SiO₂ models with different supercell sizes: $L = 2.14$ nm ($N = 648$, yellow), $L = 1.42$ nm ($N = 192$, green), and $L = 1.03$ nm ($N = 72$, red). The smallest cell exhibits a spurious peak between 7 and 8 eV, which is absent in larger systems. While the fine structure of the main peak around 10 eV varies slightly due to structural differences, the band gap remains unchanged for the $L = 1.42$ nm and $L = 2.14$ nm cells. This indicates that a box size of $L = 1.42$ nm is sufficient to avoid spurious interactions between periodic images when evaluating optical properties. Adapted from Ref. [213].

Three replicates of the 192-atoms supercell were generated (Fig. 5.4). Fig. 5.8 shows the real and

imaginary parts of $\epsilon(\omega)$ and the refractive index, which lies between 1.5 and 1.7 in the energy range of interest (8-10 eV), for the the three defect-free amorphous structures. These three structures serve as starting configurations, to which representative defects are added. Optical

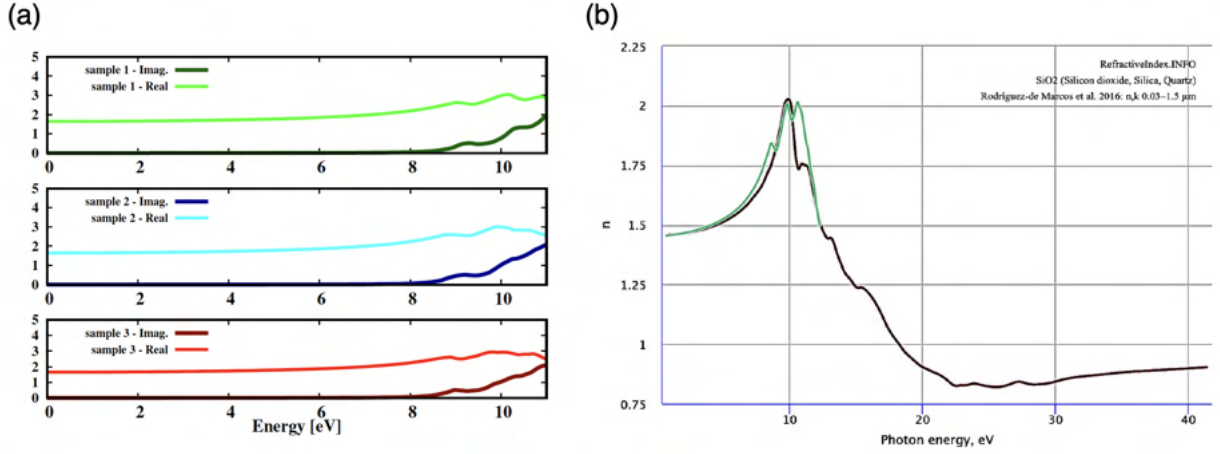


Figure 5.8: (a) Real and imaginary part of the dielectric function for the three silicon dioxide samples in Fig. 5.4, calculated with the exchange-correlation functional HSE06. (b) Refractive index of the silica sample 1 (green line) in Fig. 5.4, calculated with the exchange-correlation functional HSE06 in comparison to experimental data (black line). Adapted from Ref. [213].

properties of such defective configurations were subsequently evaluated within density functional theory using the HSE06 hybrid exchange–correlation functional [217]. The use of a hybrid functional is essential to obtain an accurate description of the electronic structure, in particular the band gap, which directly affects the calculated absorption spectra and, consequently, the imaginary part of the dielectric function.

5.5.2 Defects Play a Major Role in Optical Response Compared to the APT Static Field

The atomic evaporation process in APT occurs under the combined influence of a strong static electric field and an external laser field. To isolate and understand the role of the static field, its effect on the optical absorption properties of silica matrices was first investigated. A constant electric field of 20 V/nm was applied along the horizontal direction. This field strength is comparable to experimentally estimated values derived from the charge-state ratio (CSR) in APT analyses.

The optical properties for each configuration were computed using DFT, first within the framework of the Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional. The resulting absorption spectra were analyzed with a consistent reference energy scale, where the zero

of the energy axis is aligned with the top of the valence band for all configurations. This choice of energy reference facilitates direct comparison between different systems and highlights field-induced modifications in the electronic structure and optical response.

A systematic red shift of the optical absorption spectrum is observed in all three samples when compared to the zero electrostatic field case (see Fig. 5.9 (a)). This shift results in a slight reduction of the band gap in amorphous silica. However, the magnitude of the shift remains relatively small and does not significantly alter the overall absorption characteristics of the material. These findings suggest that the applied static electric field does not fundamentally modify the optical response of amorphous silica, but instead induces subtle and systematic perturbations in its electronic structure.

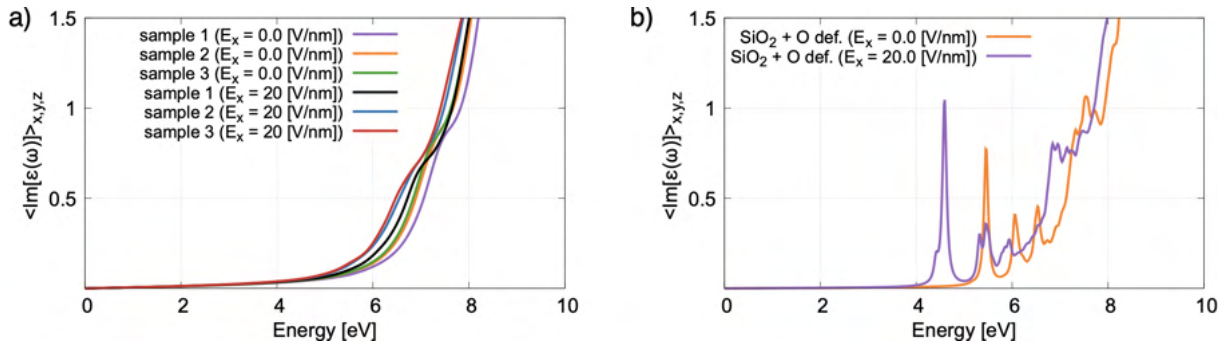


Figure 5.9: (a) Influence of the static electric field: average of the imaginary part of the dielectric tensor along the orthogonal Cartesian directions for the three SiO₂ samples. Pristine amorphous silica matrix without electric field (violet, orange and green lines) and with a constant electric field of 20 V/nm along the x direction (black, blue and red lines). (b) Amorphous silica matrix with a non-bridging oxygen defect without electric field (orange line) and with a constant electric field of 20 V/nm along the x direction (violet line). These calculations were performed with a PBE functional that notoriously underestimates the fundamental gap. Adapted from Ref. [213].

In contrast, the presence of structural defects leads to a much more pronounced effect. Specifically, when a bridging oxygen atom is removed from the amorphous silica network, a red shift of approximately 1 eV is observed in the absorption spectrum (violet line in Fig. 5.9 (b)), which is significantly larger than in the defect-free case (orange line in Fig. 5.9 (b)). This behavior is expected, as defects introduce localized intra-gap states, enhancing low-energy optical transitions that can be captured within the absorption response. Therefore, while the static electric field alone has a limited impact, its combination with structural defects plays a critical role in tuning the optical properties of amorphous silica.

5.5.3 Optical Response of Typical Defects Present in APT Analyses

A range of representative defects for APT was studied. Such impurities are known to form during sol–gel synthesis and specimen preparation processes (e.g., lift-out FIB–SEM), and may significantly influence laser light absorption. The defects considered include interstitial C, H, O, and Na atoms, as well as both interstitial and substitutional gallium atoms, in configurations either with or without nearby oxygen vacancies. The optical properties of each defective structure were computed using density functional theory (DFT) with the HSE06 exchange–correlation functional, which provides a significantly improved description of the band gap of amorphous silica compared to PBE, as reported in Fig. 5.5.

Carbon Defects

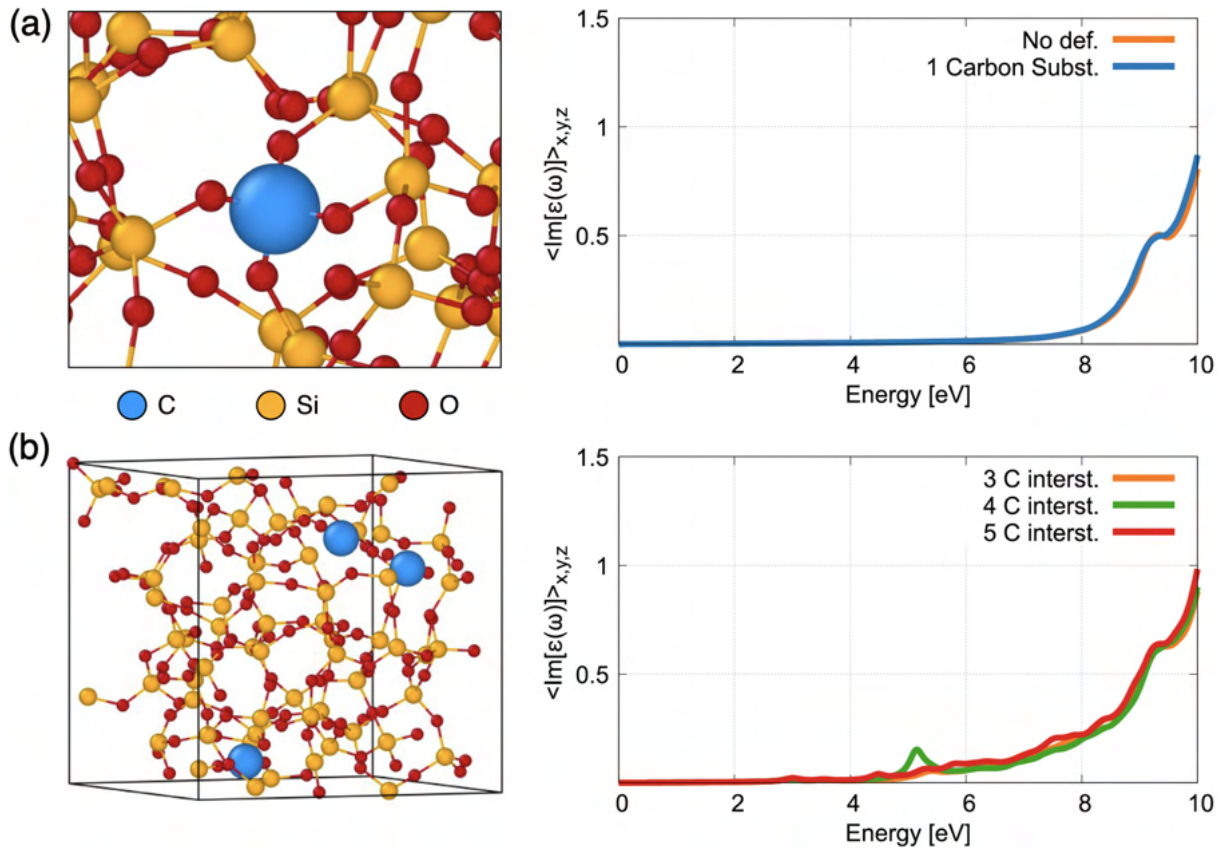


Figure 5.10: (a) Absorption spectra (right panel) of the silica sample shown in the left panel for one substitutional carbon atom (in blue color) replacing silicon (yellow color); (b) absorption spectra (right panel) of the silica samples with three (as shown in the left panel) to five interstitial carbon atoms (blue color in the left panel of the figure). Adapted from Ref. [213].

Fig. 5.10 (a) presents the calculated absorption spectra for silica structures containing substitutional carbon defects, generated by replacing a silicon atom with a carbon atom (blue atom shown in the left panel). Various configurations and amorphous environments were explored, yielding consistent results without significant variations. Fig. 5.10 (b) shows the absorption spectra for systems containing multiple interstitial carbon atoms (three to five atoms inserted into the silica network, shown in blue in the left panel). These configurations represent higher defect concentrations and varying local bonding environments.

Fig. 5.10 (a) reveals that substitutional carbon defects have a negligible impact on the optical absorption properties of silica. This behavior can be understood from the electronic structure of amorphous SiO_2 , where the valence band maximum is primarily composed of nonbonding O $2p$ orbitals. Consequently, replacing silicon with carbon does not significantly perturb the band structure or alter the band gap, in agreement with the computed absorption spectra. In contrast, interstitial carbon defects lead to a substantial increase in intragap optical absorption, particularly in the energy range between 4 and 8 eV. This enhancement arises from the introduction of localized defect states within the band gap, which facilitate additional electronic transitions at lower photon energies.

These results suggest that while substitutional defects are relatively benign, interstitial carbon impurities—especially at higher concentrations—play a critical role in enhancing light absorption in the deep-UV region. This provides a plausible explanation for the increased absorption observed experimentally.

Oxygen and Hydrogen Defects

The shape of the optical absorption spectrum is significantly altered by the introduction of specific point defects, such as an interstitial hydrogen atom bonded to oxygen or a bridging oxygen vacancy, across different configurations of the three representative silica samples (see Fig. 5.11). In the case of interstitial hydrogen, the formation of O–H bonds leads to the emergence of multiple but modest intragap absorption peaks in the energy range between 4 and 8 eV (see Fig. 5.11 (a) and (b)). Similarly, the creation of a bridging oxygen vacancy gives rise to distinct intragap features, although these are shifted toward higher energies, typically between 6 and 8 eV (see Fig. 5.11 (c)). As a result, the presence of such defects effectively reduces the optical gap and enhances sub-band-gap absorption. This behavior is consistent with experimental observations, where a reduction of the apparent band gap and increased absorption in the ultraviolet and deep-ultraviolet regions are reported. Therefore, these results highlight the critical role of hydrogen-related defects and oxygen vacancies in governing the optical response of defective silica.

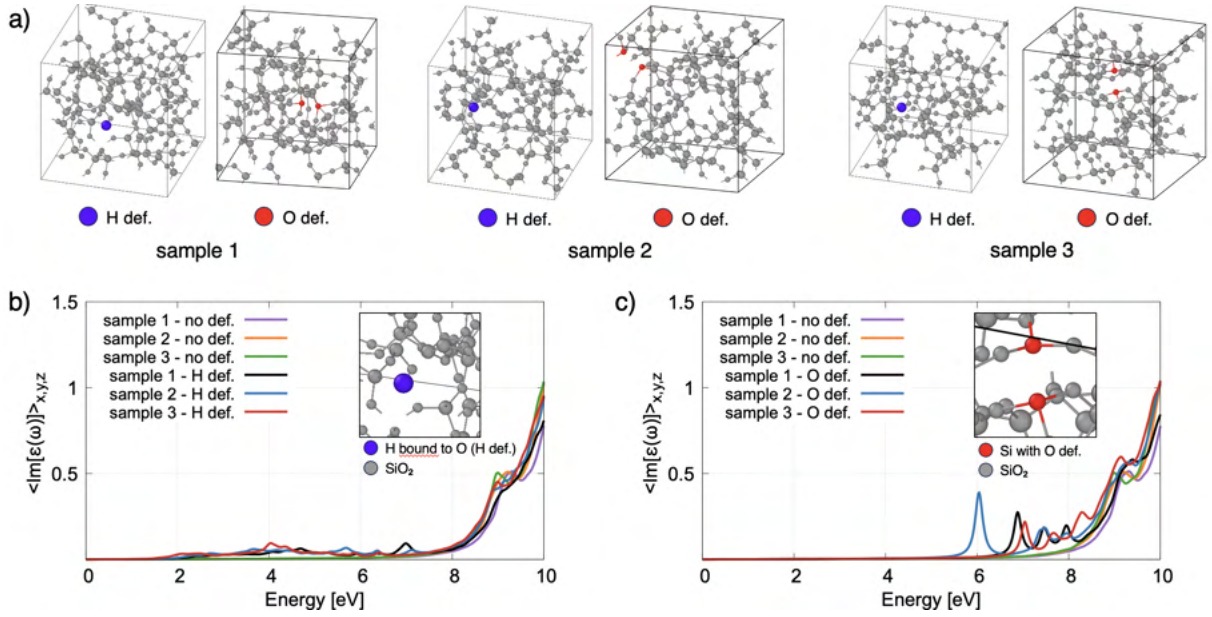


Figure 5.11: Absorption spectra of the silica samples shown in panel (a), for (b) interstitial hydrogen (H def., blue color) bound to oxygen and (c) for an oxygen vacancy (O def., red color) in different configurations compared to the defect-free amorphous structures of Fig. 5.9 (no def.). Adapted from Ref. [213].

Sodium Defects in Relation to Oxygen Vacancies

Another investigation was carried out, concerning the potential impact of sodium species in relation to SiO_2 configurations involving oxygen vacancies. Previous atomistic simulations have provided detailed insight into the structure of sodium silicate glasses, showing that sodium atoms tend to cluster in alkali-rich regions and preferentially associate with nonbridging oxygen atoms [225, 226]. To capture these effects, two representative configurations were modeled: (i) a sodium molecule embedded within the silica matrix to mimic clustering in alkali-rich regions (see left panel of Fig. 5.12 (a)), and (ii) a sodium molecule positioned near oxygen vacancies to represent aggregation around non-bridging oxygen sites (see left panel of Fig. 5.12 (b)).

The results in the right panel of Fig. 5.12 (a) indicate that the presence of interstitial atomic (blue, green) or molecular (red) sodium alone does not significantly modify the optical absorption spectra of the silica matrix. In particular, no defect-induced states are observed within the band gap (4-8 eV). Instead, the resulting spectra can be interpreted as a superposition of the contributions from isolated atomic sodium (blue and green curves), molecular sodium (red curve), and pristine silica (orange curve). In contrast, when bridging oxygen vacancies are introduced (see left panel of Fig. 5.12 (b)), sodium atoms tend to aggregate in the vicinity of

these defects. This leads to a noticeable increase in optical absorption in the energy range of approximately 4–6 eV (see right panel of Fig. 5.12 (b)). This energy range is consistent with experimental observations obtained under deep-UV irradiation, typically around 4.8 eV.

These findings suggest that while sodium alone has a negligible effect on the optical properties of silica, its interaction with intrinsic defects such as oxygen vacancies can enhance sub-band-gap absorption, thereby contributing to the experimentally observed optical response.

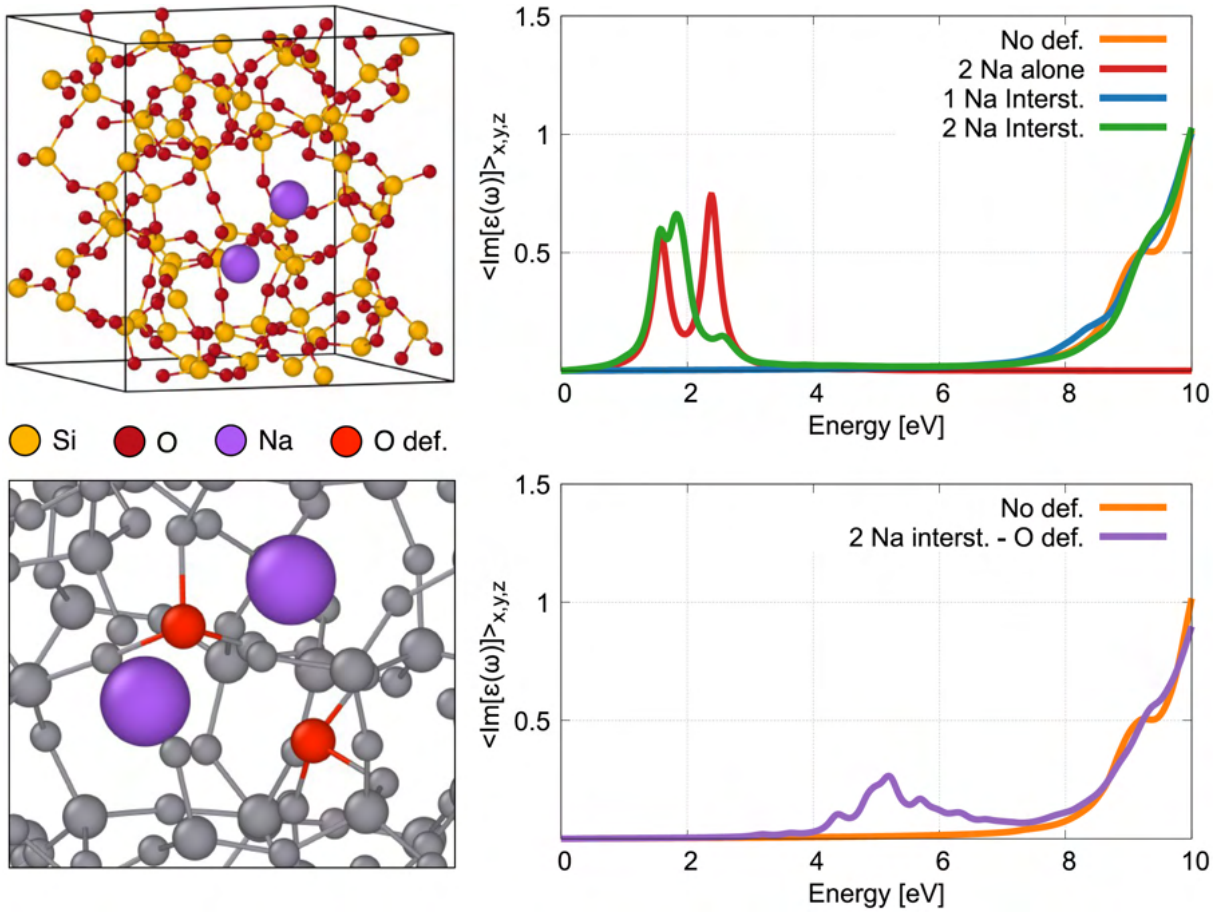


Figure 5.12: (a) Left: optimized configuration for interstitial sodium defects (violet atoms) in the amorphous silica matrix. Right: absorption spectra of the silica samples without (orange line) and with (blue and green lines) the interstitial sodium atoms, or with molecular sodium (red line). (b) Left: optimized configuration for molecular sodium near a bridging oxygen vacancy inside the amorphous silica matrix. Right: Absorption spectrum of the amorphous silica sample in the simultaneous presence of a sodium molecule in the vicinity of a bridging oxygen vacancy (violet line) and in pristine (defect-free) form (orange line). Adapted from Ref. [213].

Gallium Defects in Relation to Oxygen Vacancies

Finally, the optical absorption properties induced by the presence of gallium impurities in the silica matrix were tested, considering both interstitial (green atoms in Fig. 5.13 (a)) and substitutional (blue atoms in Fig. 5.13 (b)) configurations, as well as their interaction with bridging oxygen vacancies. Simulations reveal that substitutional incorporation of gallium leads to the

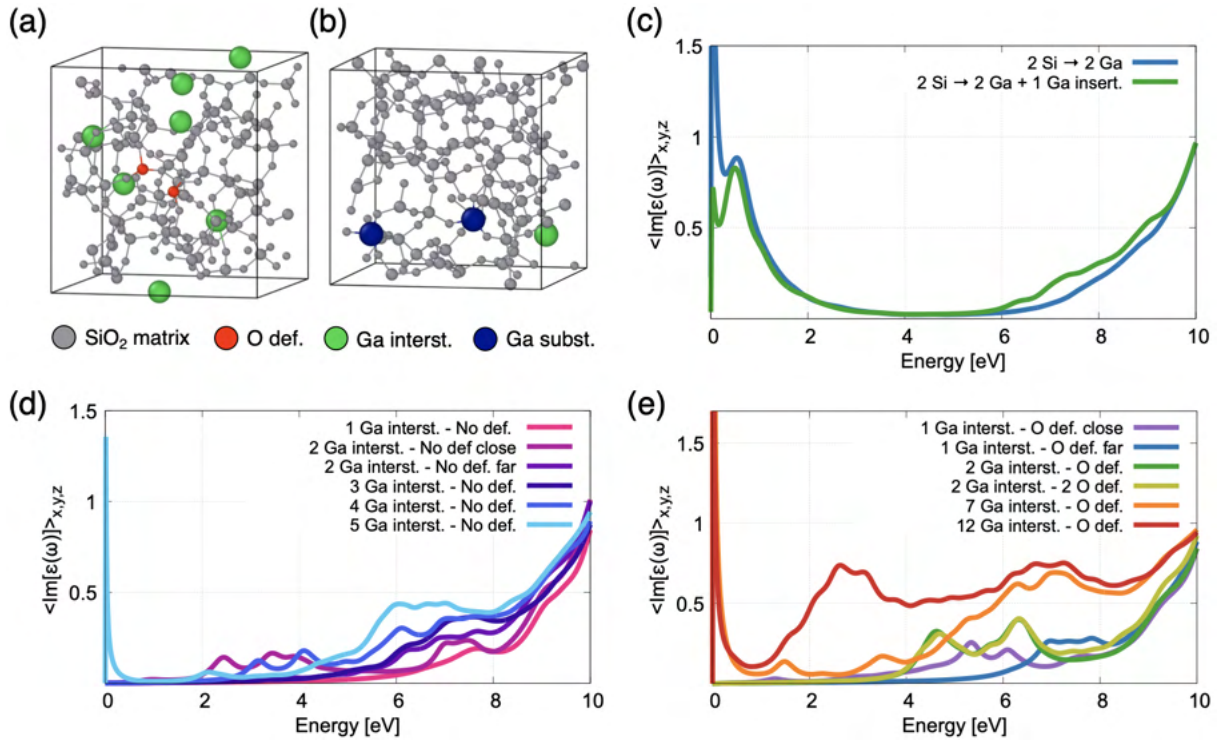


Figure 5.13: (a) Model of a silica sample with the simultaneous presence of several interstitial gallium atoms (green), one and two oxygen vacancies (red); (b) model of a silica sample with the simultaneous presence of two substitutional (blue) and one interstitial (green) gallium atom. (c) Absorption spectra of silica samples with two substitutional gallium atoms (blue line) also in the presence of interstitial gallium atoms (green line). (d) Absorption spectra of different silica samples with different numbers of interstitial gallium atoms without oxygen defects, at least in the vicinity. (e) Absorption spectra of different silica samples with different numbers of interstitial gallium atoms, which may be close to or far from the oxygen vacancy. Adapted from Ref. [213].

emergence of metallic states at the Fermi level (see green and blue features around 0 eV in Fig. 5.13 (c)). This results in a conductive electronic structure, which is in clear contradiction with experimental observations where gallium-doped silica remains insulating. Therefore, a purely substitutional model for gallium incorporation is unlikely to accurately represent the atomic-scale configuration in experimentally studied samples. In contrast, interstitial gallium

defects introduce localized intragap electronic states (see Fig. 5.13 (d)), which enhance optical absorption, particularly at low photon energies. However, the concentration of gallium plays a critical role: when the number of interstitial Ga atoms exceeds five within the simulation cell (corresponding to a concentration of approximately 2%), metallic states emerge at the Fermi level (see cyan curve in Fig. 5.13 (d)), again leading to a conductive behavior inconsistent with experiments.

At lower concentrations, specifically for four interstitial gallium atoms, the material remains non-metallic while exhibiting a significant increase in optical absorption starting from approximately 3.5 eV (see blue curve in Fig. 5.13 (d)), even in the absence of oxygen-related defects. This behavior provides a plausible explanation for experimentally observed absorption features in the UV and deep-UV spectral regions. These results indicate that interstitial gallium defects can modify the local electronic structure by introducing defect states within the band gap without necessarily inducing full metallization, depending on their concentration and spatial distribution within the host matrix.

Furthermore, the combined effect of interstitial gallium atoms and bridging oxygen vacancies was examined. Fig. 5.13 (e) shows the resulting absorption spectra for configurations with one (O def.) or two (2 O def.) oxygen vacancies, either located in close proximity to (O def. close) or far from (O def. far) the interstitial gallium atoms, for gallium concentrations up to 6%. The results demonstrate that the interplay between gallium defects and oxygen vacancies further enhances intragap absorption.

Overall, these findings suggest that the coexistence of interstitial gallium atoms and oxygen vacancies in various configurations might contribute to the experimentally observed optical absorption in the UV and deep-UV range, together with the other tested defects.

5.6 Conclusive Remarks

The present study provides a comprehensive investigation of the optical absorption properties of sol-gel silica in the context of laser-assisted atom probe tomography (La-APT) under green, ultraviolet (UV), and deep-ultraviolet (DUV) illumination. The results demonstrate that the improved performance of UV and DUV La-APT, compared to green illumination, originates from the enhanced optical absorption of silica at shorter wavelengths, primarily induced by the presence of intrinsic and extrinsic defects.

Defects such as oxygen vacancies, as well as interstitial impurities introduced during synthesis and sample preparation (e.g., C, H, Na, and Ga), give rise to localized electronic states within the band gap. These states enable sub-bandgap optical absorption, particularly in the UV and DUV energy ranges. Increased absorption leads to a higher generation of free charge carriers, promoting more efficient energy transfer to the lattice and resulting in enhanced local heating of the silica matrix. Consequently, lower external electric fields are required to trigger field evaporation, thereby reducing mechanical stress and minimizing the risk of specimen fracture.

Additionally, the increase in free carrier density reduces the effective resistivity of the material, mitigating potential drops along the tip and enabling a higher effective electric field at the apex. This facilitates more stable evaporation conditions. Experimental absorption spectra confirm a significant increase in absorption at wavelengths below 300 nm, with an enhancement by nearly a factor of three between 343 nm and 258 nm [213], consistent with the presence of defect-induced sub-bandgap states.

Density functional theory (DFT) calculations using the HSE06 hybrid functional corroborate these findings. This thorough defect analysis has highlighted four particularly relevant cases: (i) the oxygen vacancy that disrupts the Si–O–Si network by breaking a Si–Si bond; (ii) the interstitial hydrogen defect that passivates and weakens the oxygen bonds; (iii) an oxygen vacancy with a Na₂ molecule nearby; (iv) interstitial and substitutional gallium atoms, which could originate from the synthesis process of the needle-shaped silica specimens for the APT analyses, also in the presence of bridging oxygen vacancies. It was shown that these specific defects—particularly interstitial carbon, hydrogen-related defects, oxygen vacancies, and sodium in the presence of vacancies—enhance absorption in the DUV region, while having a limited effect in the UV range. The computed spectra exhibit very low absorption at photon energies around 3.6 eV, even at high defect concentrations, supporting the experimentally observed lower efficiency of UV illumination compared to DUV.

The role of gallium impurities, potentially introduced during focused ion beam (FIB) preparation, has also been examined. Substitutional gallium and high concentrations of interstitial gallium lead to the formation of metallic states at the Fermi level, which is inconsistent with the experimentally observed insulating behavior of silica. In contrast, moderate concentrations

of interstitial gallium (on the order of 2%), particularly in the presence of oxygen vacancies, introduce intragap states that enhance absorption in the 3.5–5 eV range without inducing metallization. These configurations might provide a possible explanation for experimentally observed absorption features.

The influence of strong static electric fields, on the order of 20 V/nm, has also been assessed. The results indicate that such fields induce only minor modifications in the optical absorption spectrum, suggesting that the optical response of amorphous silica is robust under APT-relevant field conditions.

Finally, although the present DFT approach captures the essential role of defect-induced absorption, further theoretical improvements—such as the inclusion of excitonic effects via many-body perturbation theory (e.g., Bethe–Salpeter equation)—would provide a more accurate description of the optical response. However, such methods remain computationally demanding for systems of this size. Future work should therefore focus on more complex defect configurations and multiscale approaches to fully elucidate the interplay between defects, optical absorption, and evaporation mechanisms in silica-based systems.

My PhD research was conducted within the broader framework of the MIMOSA project, a large-scale European-funded consortium that brings together numerous research groups across several European countries. The consortium is characterized by its interdisciplinary nature, encompassing expertise ranging from experimental investigations to theoretical modeling. During the course of my PhD, I had the opportunity to collaborate with two experimental research groups: the group led by A. Vella at the University of Rouen (France), and the group led by M. Andersson at Chalmers University of Technology in Gothenburg (Sweden). These collaborations provided valuable experimental perspectives that complemented the theoretical aspects of my work.

Atom Probe Tomography (APT) has recently emerged as a promising technique for the structural characterization of biological systems at the single-molecule level. Traditionally developed for conductive and semi-conductive materials, APT faces intrinsic challenges when applied to biomolecules, primarily due to their fragility, low electrical conductivity, and sensitivity to external fields and thermal perturbations. Early explorations into bio-APT demonstrated that embedding biological specimens within a stabilizing matrix—most notably amorphous silica—can mitigate these limitations by providing mechanical support, chemical protection, and improved field evaporation conditions. This strategy has opened a pathway toward extending APT into the realm of structural biology, with the long-term goal of achieving near-atomic resolution reconstructions of biomolecular systems while preserving their native conformational landscape.

Within this context, the three computational studies considered here—in Chapters 3, 4 and 5—collectively establish a coherent and multiscale understanding of the key processes governing silica-assisted APT of biological materials. Rather than addressing isolated aspects, they converge toward a unified picture in which the success of APT in biological applications depends critically on the interplay between silica chemistry, electronic excitation mechanisms, and field-driven evaporation dynamics. Their conclusions, when integrated, define both the current capabilities and the fundamental constraints of the technique.

A central conclusion emerging from the first study is that silica encapsulation is not a passive

embedding process but an active, highly selective interaction between silica species and biomolecular surfaces. The speciation of silica –whether monomeric or oligomeric– plays a relevant role in determining how the matrix forms and stabilizes around the target molecule. Monomeric species behave largely as a homogeneous solvent component, exerting minimal structural influence. In contrast, oligomeric species exhibit a strong propensity to accumulate at the biomolecular interface, displacing water and forming structured, persistent adsorption layers. This distinction leads to a key conceptual advancement: the early stages of silica matrix formation are governed by cooperative and localized interactions rather than uniform condensation. The resulting matrix is therefore inherently heterogeneous at the nanoscale, with regions of enhanced silica density correlating with specific structural features of the biomolecule. In particular, flexible and intrinsically disordered regions emerge as preferential nucleation sites for silica adsorption. Their conformational plasticity and exposure of hydrophilic and charged residues create energetically favorable conditions for interaction with oligomeric silica species. From a structural perspective, this selective adsorption has a dual implication. On one hand, it preserves the integrity of well-folded domains, confirming that silica embedding can maintain native-like conformations in structured regions. On the other hand, it actively modulates the dynamics of disordered segments, reducing their conformational freedom and, in some cases, stabilizing transient secondary structures like α -helices. The conclusion is therefore not merely that APT-compatible embedding is feasible, but that it introduces a controlled perturbation of the conformational ensemble – one that can potentially be exploited to probe otherwise elusive structural states.

The second study extends this understanding into the field evaporation regime, addressing one of the most critical bottlenecks in APT of insulating materials: the efficient and controlled removal of ions under extreme electric fields. The key conclusion here is that ion evaporation from insulating silica cannot be explained solely in terms of static electric field strength. Instead, it is governed by a complex coupling between ionization (charge transport), electronic dynamics, and transient field parameters/conditions. The introduction of single-cycle, high-intensity Terahertz (THz) pulses reveals a fundamentally new evaporation mechanism in the context of SiO_2 matrices. The observed polarity dependence –where negative pulses dramatically enhance evaporation efficiency while positive pulses suppress it– demonstrates that the temporal structure and directionality of the electric field are as important as its magnitude. This resulting asymmetry is not a trivial experimental artifact but a direct manifestation of the underlying electron dynamics in an insulating environment. The theoretical models developed in this study lead to a decisive conclusion: electron emission and subsequent electron dynamics near the tip apex are the primary drivers of evaporation in these systems. In particular, negative pulses induce conditions that favor electron extraction followed by partial backscattering toward the tip. This process results in localized energy deposition, effectively generating nanoscale thermal

hotspots. These transient heating events lower the activation barrier for cation evaporation, assisting and enabling efficient ion emission even in materials with poor intrinsic conductivity. This mechanism redefines the role of electronic processes in APT. Rather than acting merely as a background facilitator of field evaporation, electron dynamics becomes the central element controlling evaporation efficiency. The implication is profound: by tailoring the temporal and polarity properties of the THz single-cycle pulse, it is possible to manipulate evaporation pathways in a controlled manner. Moreover, the conclusion that such polarity-dependent effects could not be restricted to the THz regime suggests a broader applicability of this mechanism across different excitation conditions, provided that sufficient electron emission is achieved.

The third study completes the picture by addressing the optical properties of silica under laser-assisted APT (la-APT) conditions. Its primary conclusion is that the efficiency of laser-assisted evaporation in the APT amorphous SiO_2 tip is fundamentally limited –and simultaneously enabled– by defect-mediated optical absorption. Pure silica has a very large and wide band gap (circa 9 eV). However, the presence of intrinsic and extrinsic defects introduces localized electronic states within the band gap, enabling sub-bandgap absorption and thereby facilitating energy transfer from the laser pulse to the material. This leads to a crucial insight: defects are not merely unavoidable imperfections but functional components that determine the operability of the technique. Their presence enhances the generation of free charge carriers, increases local heating, and reduces the effective electric field required for evaporation. Consequently, defect engineering becomes a central parameter in optimizing APT performance. The study further concludes that not all defects contribute equally. Specific configurations –such as oxygen vacancies, hydrogen-related defects, and impurity complexes involving light elements– are particularly effective in enhancing deep-UV absorption without compromising the insulating nature of the material. In contrast, excessive incorporation of certain species, such as gallium, can induce metallic behavior, which is detrimental to controlled evaporation. This establishes a delicate balance between enhancing optical absorption and preserving the electrical characteristics necessary for stable APT operation. An additional important conclusion is that the optical response is largely independent of the high-field environment characteristic of APT, but only when defects are not introduced.

When considered together, these three studies converge toward a unified framework for silica-assisted APT of biological systems. The encapsulation process defines the initial structural state of the biomolecule within a heterogeneous and functionally active matrix. The optical properties of this matrix, governed by its defect landscape, determine how efficiently external energy can be coupled into the system. Finally, the interplay between electronic excitation and transient electric fields dictates the mechanisms by which ions are ultimately evaporated and detected. The final conclusion is that successful application of APT to biological materials requires a

holistic optimization strategy that simultaneously addresses chemical, structural, electronic, and optical factors.

Importantly, these works collectively suggest that the limitations traditionally associated with APT of insulating and biological materials are not fundamental barriers but challenges that can be overcome through targeted control of material properties and excitation conditions. Silica is not simply a protective medium but a multifunctional component that mediates every stage of the experiment –from sample preparation to ion detection. Thus, to recall, by tuning silica speciation during embedding, engineering defect populations to optimize optical absorption, and exploiting advanced pulsed-field techniques to control electron dynamics, it becomes possible to achieve stable, high-resolution measurements while preserving the essential structural features of the biomolecule.

In this light, APT could emerge in the future not only as a viable tool for biological analysis but as a technique with unique capabilities. Its sensitivity to both atomic composition and local environment, combined with the ability to probe individual molecules, positions it as a complementary approach to established structural biology methods. Looking forward, these studies described in this thesis suggest several key directions. Optimization of key steps such as silica polymerization, evaporation conditions and meaningful defect incorporation might lead to large improvements in APT pipelines. Moreover, advanced theoretical methods (for instance, Machine Learning techniques) that might bridge electronic and atomistic levels of detail, will further refine our understanding of the fundamental APT mechanisms. Finally, systematic experimental validation across a broader range of biomolecular systems will be necessary to fully establish the generality of the observed phenomena, that for the moment are just somehow confined to a “proof of concept” framework.

In conclusion, the combined findings of these studies –having a dual nature, computational and experimental– define a coherent and predictive framework, that leads to an integrated perspective providing a solid foundation for the development of APT as a powerful tool for single-molecule structural analysis.

Other Projects

In parallel with the main research activities carried out within the Statistical and Biological Physics Group, my PhD work has also involved several collaborative projects that either originated before my enrollment in the doctoral program or were initiated during its course. Although these studies addressed different biological questions, they shared a common methodological framework: the investigation of specific protein–ligand interactions through computational molecular modeling and simulations, with the aim of complementing and rationalizing experimental observations obtained by the collaborating research groups. The collaborations included:

- the Laboratory of Translational Defects of Gene Expression at the Institute of Biophysics (CNR-IBF, Trento), led by Dr. Gabriella Viero. This project investigates the interaction between a known small-molecule ligand and a specific protein, implicated in the pathophysiology of spinal muscular atrophy (SMA). The study aims to elucidate the molecular basis of such interaction and its potential role, by flanking the most relevant experimental outcomes with molecular docking and molecular dynamics simulations.
- the Armenise-Harvard Laboratory of Cell Division within the Department of Cellular, Computational and Integrative Biology of the University of Trento, led by Prof. Luca Fava. This collaboration concerns the study of N⁶-methyladenosine (m⁶A), the most abundant internal RNA modification in eukaryotic messenger RNA. The deposition of m⁶A is catalyzed by the methyltransferase complex composed primarily of METTL3 and METTL14. Aberrant m⁶A methylation has been implicated in numerous human diseases, particularly cancer. Consequently, the METTL3–METTL14 complex has emerged as an attractive therapeutic target, promoting the development of several small-molecule inhibitors aimed at modulating its methyltransferase activity. Although these inhibitors have shown promising preclinical results, the potential emergence of resistance-conferring mutations in METTL3 or METTL14 represents a possible challenge for their future clinical application. Such mutations could reduce inhibitor binding while preserving enzymatic function, ultimately compromising therapeutic efficacy. This project aims at identifying

the molecular mechanisms underlying the interaction between a subset of drugs and clinically relevant mutants, through an integrated computational and experimental approach. In particular, I employed –and am currently using– molecular dynamics simulations and enhanced sampling techniques based on free-energy calculations, including metadynamics, to characterize the effects of several mutations on ligand binding and conformational dynamics.

- the Bio-organic Chemistry Laboratory and the Laboratory of Protein Crystallography and Structure-Based Drug Design at the University of Trento, specifically Dr. Jacopo Vigna and Prof. Graziano Lolli. Specifically, the aim of this study was to identify and characterize small-molecule inhibitors of YTHDF proteins capable of disrupting the recognition of m⁶A-modified RNA. Particular emphasis was placed on elucidating the non-covalent binding mechanism of ebselen and its structural effects on the YTH domain to provide a framework for the rational design of improved YTHDF inhibitors. In this very broad context, my role was to support with molecular dynamics simulations and enhanced methods (mainly well-tempered metadynamics and localized volume-based metadynamics) the most informative experimental outcomes.

The preliminary results of the first two collaborations obtained so far are currently subject to intellectual property protection and some are also undergoing patent procedures. Consequently, their scientific findings cannot be disclosed yet. Finally, the third collaboration has not reached a sufficiently advanced stage to be published, yet. However, I decided to insert a comprehensive description of its scientific rationale and computational methodology in Appendix C. The corresponding molecular simulations and data analyses are currently underway, and the complete results are expected to become available in the near future.

Appendices

A.1 Supplementary Figures

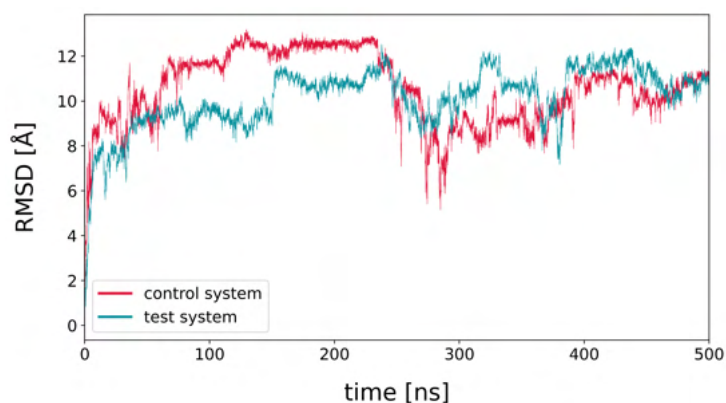


Figure A.1: RMSD of the *control* (red) and *test* (blue) trajectories of the transmembrane domain of IL22RA1. Clearly, the helix does not converge to a specific structure at these simulation timescales in both scenarios. Adapted from Ref. [115].

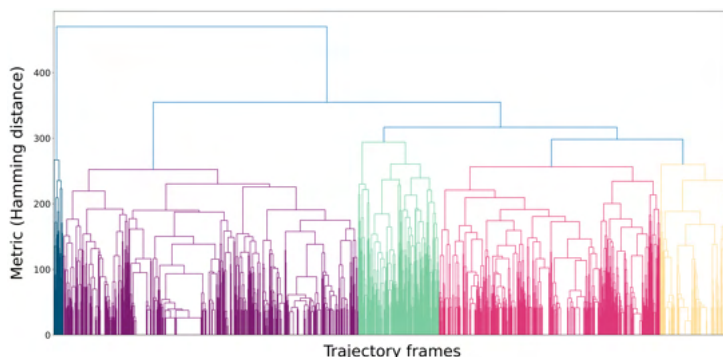


Figure A.2: Dendrogram of the hierarchical clustering analysis on the merged (*control*, *test*) trajectory of the transmembrane helix of IL22RA1, whereby five clusters were isolated based on the dynamical evolution of the secondary structure elements. Adapted from Ref. [115].

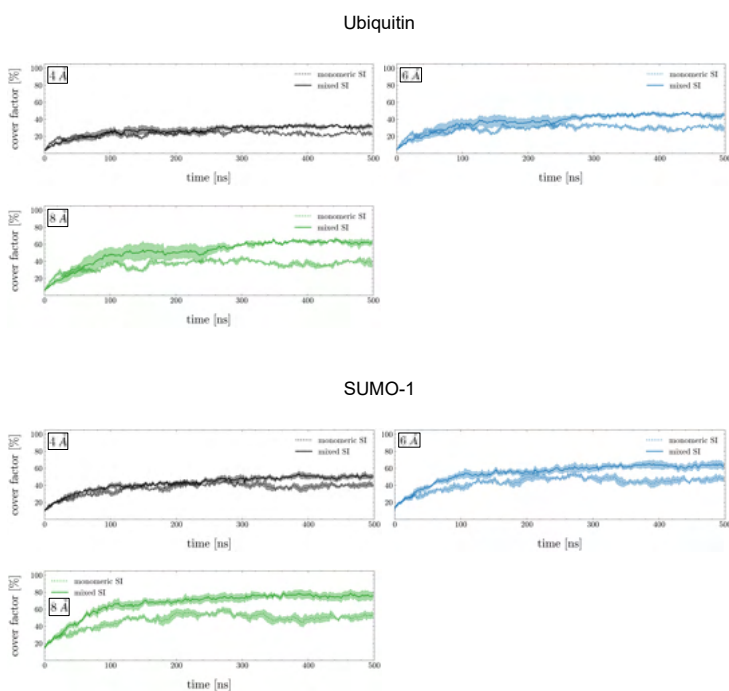


Figure A.3: Covering factor of Ubiquitin and SUMO-1 by silica precursors - i.e., the fraction of silicon atoms within 4 Å (black), 6 Å (blue), 8 Å (green) of the protein surface (data are averaged over MD replicates). The monomeric and mixed scenarios are shown as dotted and solid lines respectively. Threshold distance criteria are based on the RDFs shown in Fig. 3.9. Results are averaged among the three replicates for each studied system. Adapted from Ref. [116].

A.2 Simulation Protocol for the Reproduction of the Dupuis *et al.* Results

The ReaxFF simulations in this thesis follow the protocol established by Dupuis [173], in which a hybrid grand canonical Monte Carlo (GCMC) and parallel tempering (PT) approach is used to enhance sampling of silica gel structures.

A.2.1 Software and Initial Configuration

All simulations were performed using the LAMMPS molecular simulation package. The initial atomic coordinates were taken from and consisted of a 100 $\text{Si}(\text{OH})_4$ molecules placed into a water box of 35-30-15 Å; periodic boundary conditions were applied in all directions.

A.2.2 Force Field

Atomic interactions were described using the Fogarty *et al.* ReaxFF model [175].

A.2.3 Energy Minimization and Equilibration

Prior to production sampling, the system was energy-minimized using the LAMMPS implementation of the Steepest Descent algorithm [179]. Equilibration consisted of a canonical ensemble (NVT) run for 100 ps with a timestep of 0.1 fs at $T = 298$ K using the Langevin thermostat implementation of LAMMPS [179]. For such small equilibration times and without PT or extremely high temperatures, chemical reactions among the $\text{Si}(\text{OH})_4$ molecules are extremely rare, so there is no need to worry about silica polymerization.

A.2.4 Production Simulation

Production sampling was conducted for 1 ns under the same NVT conditions of equilibration, with the same timestep of 0.1 fs, with PT and GCMC implementations within the LAMMPS suite [179].

A.2.5 Hybrid GCMC–PT Scheme for Reactive MD

To model water exchange and structural relaxation in reactive silica systems, a hybrid simulation approach was employed, that couples GCMC with PT within a ReaxFF-based molecular dynamics (MD) framework. This methodology enables efficient sampling of both configurational and compositional phase space, particularly in systems where bond formation and breaking are relevant.

GCMC Step with MD Relaxation

Each GCMC cycle consists of a sequence of 10 trial moves, where each trial is separated by 10 fs of MD in the NVT ensemble. Unlike traditional GCMC schemes that consider only particle insertion or deletion, each trial move here includes:

- An insertion or deletion of a water molecule,
- A short MD relaxation to allow local structural rearrangement.

The acceptance or rejection of a trial move is determined using a Metropolis criterion based on the Boltzmann-weighted energy difference between the initial and final states:

$$P_{\text{acceptance}} = \min(1, \exp[-\beta(\Delta E - \mu_{\text{H}_2\text{O}}\Delta N)]), \quad (\text{A.1})$$

where $\beta = k_B T$, ΔE is the energy difference after the insertion/deletion plus relaxation, $\mu_{\text{H}_2\text{O}}$ is the chemical potential of water, and ΔN is the change in the number of water molecules. This strategy allows inserted molecules to diffuse toward energetically favorable sites before evaluation, leading to significantly higher acceptance rates compared to conventional GCMC moves, albeit at a higher computational cost.

Chemical Potential Calibration

The chemical potential of water, $\mu_{\text{H}_2\text{O}}$, is determined from separate bulk simulations using ReaxFF. A value of $\mu_{\text{H}_2\text{O}} = 225$ kcal/mol was found to reproduce a density of 1 g/cm³. Adjusting this parameter allows controlled hydration or dehydration of the silica network in the box.

Parallel Tempering Phase

Following each GCMC phase, the system undergoes a PT simulation consisting of 500 steps. A set of 16 replicas is simulated at temperatures ranging from 300 K to 1000 K. Replica exchange attempts are performed with a ratio of 1 exchange attempt every 10 MD steps. The acceptance probability for replica exchange is governed by:

$$P_{\text{swap}} = \min(1, \exp[(\beta_i - \beta_j)(E_j - E_i)]), \quad (\text{A.2})$$

where i and j denote neighboring replicas. An average exchange acceptance rate of approximately 10% ensures efficient sampling across the PT ladder and helps overcome energy barriers, such as the ~ 15 kcal/mol barrier associated with siloxane bond formation under gelation conditions [173]. At temperature T , the relevant thermal energy scale is given by $k_B T$. At 300 K:

$k_B T \approx 0.6$ kcal/mol. At 1000 K: $k_B T \approx 2.0$ kcal/mol Therefore, a barrier of 15 kcal/mol corresponds to: $\sim 25 k_B T$ at 300 K, $\sim 7-8 k_B T$ at 1000 K.

Alternating GCMC and PT Cycles

The overall simulation alternates between GCMC and short PT phases. During the GCMC stage, the number of water molecules is equilibrated according to the imposed chemical potential, while the PT phase enhances configurational sampling and facilitates structural transitions. This combined approach is particularly suited for modeling the silica gel formation, in which slow chemical reactions and solvent exchange play a critical role.

A.3 Modeling the pH in Molecular Dynamics

Modeling the effect of pH in molecular dynamics (MD) simulations is essential for accurately describing systems where protonation states play a critical role, like for example the ones studied in this thesis, composed by biomolecules and silicon–oxide interfaces. Unlike standard MD, where protonation states are fixed, constant-pH molecular dynamics (CpHMD) allows these states to fluctuate in response to the imposed pH. This section provides a comprehensive overview of the theoretical frameworks, practical implementations, and limitations of CpHMD methods.

Notably, this is something different from what described in the previous sections: here the aim is not to model reactivity, but to monitor the system dynamics according to a given pH value. For example, all the notions that are present in the following lines could be used to study how a ligand modifies the local pH within its binding pocket, stabilizing a protonation state that is not favoured by e.g. water (apo-state of the binding pocket). This section is inspired by Refs. [227, 228].

A.3.1 Fundamental Concept of Constant-pH Simulations

In constant-pH simulations, the protonation state of titratable sites is treated as a dynamic variable coupled to the thermodynamic reservoir defined by the pH. The probability of a given protonation state is governed by:

$$P \propto \exp \left[-\beta \left(E(\mathbf{r}, \mathbf{n}) + \sum_i n_i \mu_{\text{H}^+} \right) \right], \quad (\text{A.3})$$

where $\mathbf{r}$ are atomic coordinates, $\mathbf{n}$ represents protonation states, E is the potential energy, μ_{H^+} is the chemical potential of the proton. The proton chemical potential is related to pH through:

$$\mu_{\text{H}^+} = \mu_{\text{H}^+}^\circ + k_B T \ln(10) \cdot \text{pH}. \quad (\text{A.4})$$

and this equation is proportional to the energy needed to insert a proton into a molecule. In the next sections two approaches will be explained, that integrate in two different ways such concepts.

A.3.2 Discrete Constant-pH Molecular Dynamics

In the discrete approach, protonation states are treated as binary variables (protonated/deprotonated), and transitions are performed via Monte Carlo (MC) moves interleaved with MD. A trial change

in protonation state is accepted or rejected using a Metropolis criterion:

$$P_{\text{acc}} = \min(1, \exp[-\beta(\Delta E + \Delta G_{\text{pH}})]), \quad (\text{A.5})$$

where:

$$\Delta G_{\text{pH}} = k_B T \ln(10) \cdot (\text{pH} - \text{p}K_a) \cdot \Delta n. \quad (\text{A.6})$$

Here, ΔE is the energy difference between protonation states, $\Delta n = \pm 1$ is the change in proton count, $\text{p}K_a$ is the intrinsic acidity constant of the site. This scheme resembles GCMC, where proton insertion/removal is analogous to particle exchange. A possible algorithm workflow can be:

1. Run MD for a fixed number of steps.
2. Select a titratable site, e.g. amino acid within a protein. This choice can be random (i.e. given a set of titratable sites, select one every fixed number of timesteps), or sequential (every fixed number of timesteps, perform steps 3., 4. and 5. over every titratable site, one after the other).
3. Attempt protonation/deprotonation.
4. Accept/reject *via* Metropolis criterion.
5. Continue MD with updated state. If the solvent is explicit, perform a relaxation step before continuing, in order to let the system adapt to the insertion of new sets of atoms.

This approach is intuitively highly computationally expensive.

A.3.3 Continuous Constant-pH Molecular Dynamics

In continuous CpHMD, protonation is described by a continuous variable $\lambda \in [0, 1]$, interpolating between protonated and deprotonated states. The formalism of continuous CpHMD starts from a Hamiltonian re-formulation and is very close to what happens for Free Energy Perturbation methods [229, 230]:

$$H(\mathbf{r}, \lambda) = (1 - \lambda)H_{\text{prot}}(\mathbf{r}) + \lambda H_{\text{deprot}}(\mathbf{r}) + U_{\text{bias}}(\lambda), \quad (\text{A.7})$$

where U_{bias} enforces correct thermodynamics under a given pH value (see later). The λ coordinate plays a pivotal role in this formalism: the protonation state of a titratable site is not treated as a discrete variable, but rather as a continuous coordinate $\lambda \in [0, 1]$. This variable provides a smooth interpolation between two physical states:

- $\lambda = 0$: fully protonated state,
- $\lambda = 1$: fully deprotonated state.

Intermediate values ($0 < \lambda < 1$) correspond to nonphysical hybrid states, which are nevertheless useful for enhancing sampling and enabling continuous transitions. In equation A.7, the total Hamiltonian of the system depends explicitly on λ , as H_{prot} and H_{deprot} are the Hamiltonians of the protonated and deprotonated states, and $U_{\text{bias}}(\lambda)$ is a biasing potential ensuring correct thermodynamic weighting. This linear interpolation effectively mixes force field parameters (charges, bonded terms, etc.) between the two states.

The λ variable is treated as an additional dynamical degree of freedom with an associated fictitious mass m_λ . Its evolution follows:

$$m_\lambda \ddot{\lambda} = -\frac{\partial H(\mathbf{r}, \lambda)}{\partial \lambda}. \quad (\text{A.8})$$

This allows λ to evolve continuously during the simulation, enabling spontaneous transitions between protonation states.

The equilibrium distribution of λ is related to the free energy difference between protonation states. The effective free energy profile along λ is:

$$F(\lambda) = -k_B T \ln P(\lambda), \quad (\text{A.9})$$

where $P(\lambda)$ is the probability density of observing a given λ . To recover physically meaningful protonation states, one typically enforces a double-well potential:

$$U_{\text{restrain}}(\lambda) = k_\lambda \lambda^2 (1 - \lambda)^2, \quad (\text{A.10})$$

which favors $\lambda \approx 0$ and $\lambda \approx 1$. The effect of pH is introduced via a biasing term, $U_{\text{bias}}(\lambda)$, also written as:

$$U_{\text{pH}}(\lambda) = k_B T \ln(10) (\text{pH} - \text{p}K_a) \lambda, \quad (\text{A.11})$$

which shifts the relative stability of the λ states. If $\text{pH} < \text{p}K_a$, the protonated state ($\lambda \approx 0$) is favored; if $\text{pH} > \text{p}K_a$, the deprotonated state ($\lambda \approx 1$) is favored. Although intermediate λ values do not correspond to real chemical states, they:

- Lower free energy barriers between protonation states,
- Enable efficient sampling,
- Allow coupling with standard MD integrators.

Thus, λ acts as an *alchemical coordinate* that transforms one chemical state into another in a continuous manner, analogous to free energy perturbation or thermodynamic integration methods.

A.3.4 Coupling Constant-pH Methods with Replica Exchange

To enhance sampling, CpHMD can be combined with Replica Exchange Molecular Dynamics (REMD) (equivalent of Parallel Tempering), either in temperature or pH space.

In the former case (temperature space), each replica is simulated at a different temperature T_i , with exchange probability:

$$P_{\text{swap}} = \min(1, \exp[(\beta_i - \beta_j)(E_j - E_i)]). \quad (\text{A.12})$$

This improves barrier crossing in both conformational and protonation spaces. Alternatively, in the pH space implementation, replicas differ in pH values. The acceptance criterion becomes:

$$P_{\text{swap}} = \min(1, \exp[-\beta \ln(10)(\text{pH}_i - \text{pH}_j)(N_j - N_i)]), \quad (\text{A.13})$$

where N_i is the number of bound protons in replica i .

Also hybrid temperature-pH exchange schemes can be utilized. A generalized exchange criterion:

$$P_{\text{swap}} = \min(1, \exp[-(\beta_i E_i + \beta_i \mu_i N_i - \beta_j E_j - \beta_j \mu_j N_j)]). \quad (\text{A.14})$$

This allows simultaneous sampling of temperature and proton chemical potential.

A.3.5 Applicability of CpHMD to the Systems in this Thesis

Applying CpHMD to systems such as the Waterglass or the Mixed scenarios is conceptually appealing but practically challenging. Key limitations are:

- *Computational Cost:* CpHMD requires additional sampling over protonation states, significantly increasing computational expense. Think about the fact that, in principle, every silica molecule has four titrable sites, and there are also the ones coming from the protein amino acids.
- *Timescale Mismatch:* Gelation and siloxane formation involve barriers (~ 15 kcal/mol) that already require enhanced sampling (e.g., PT). Adding CpHMD further complicates convergence.

In conclusion, while constant-pH methodologies provide a rigorous framework for modeling

proton-coupled phenomena, their application to large-scale systems displaying lots of titrable sites remains limited by computational cost.

Supplementary Data of Chapter 4

B.1 Supplementary Figures

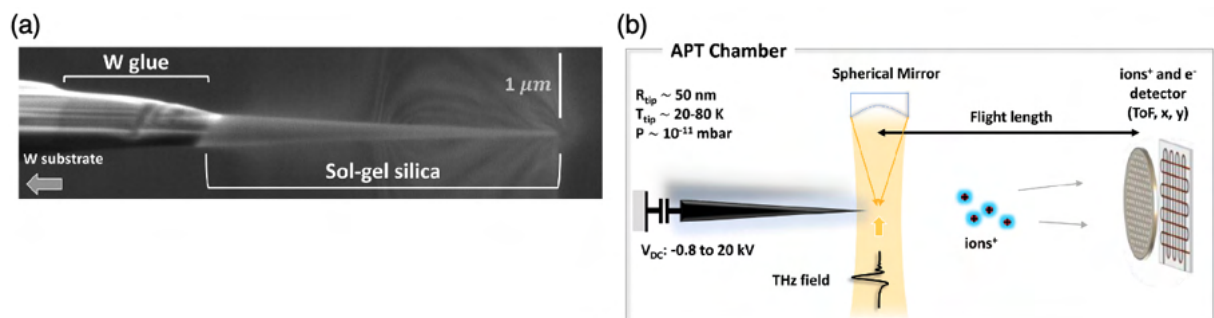


Figure B.1: APT analysis of the SiO_2 and LaB_6 samples. (a) SEM Image of the APT chamber; the needle-shaped sample is attached to a Tungsten (W) support with a Tungsten glue. (b) Scheme of the whole APT chamber. Adapted from Ref. [184].

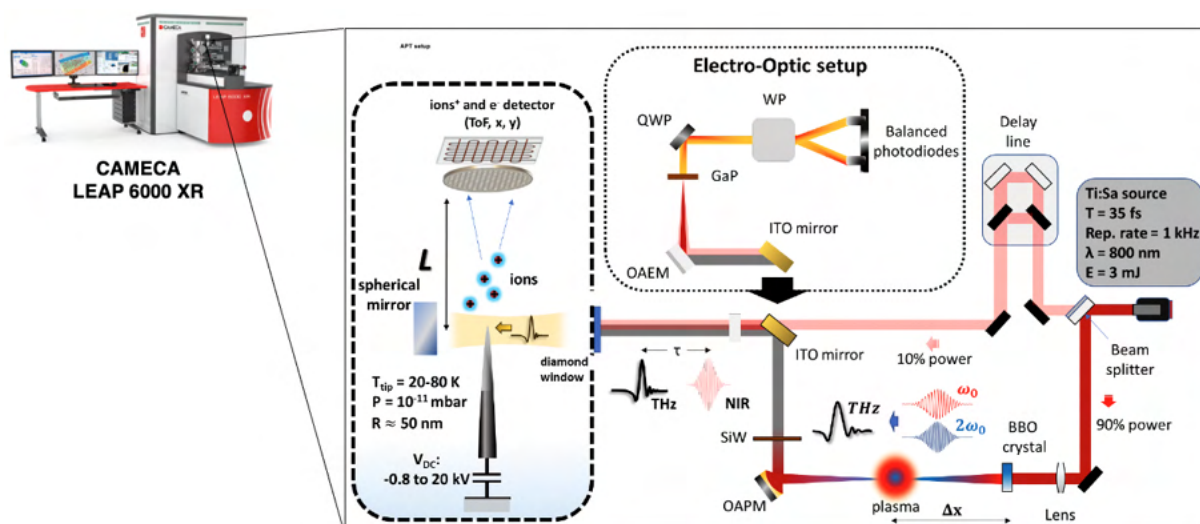


Figure B.2: Low repetition rate TERA-SAT setup (1 kHz). THz radiation generation involves a two-color plasma process where a high-energy NIR beam is focused through a lens and a BBO crystal, converting part of the fundamental wave into its second harmonic to produce a bi-chromatic femtosecond beam. This beam forms a plasma filament that emits a broadband THz pulse. Adapted from Ref. [184].

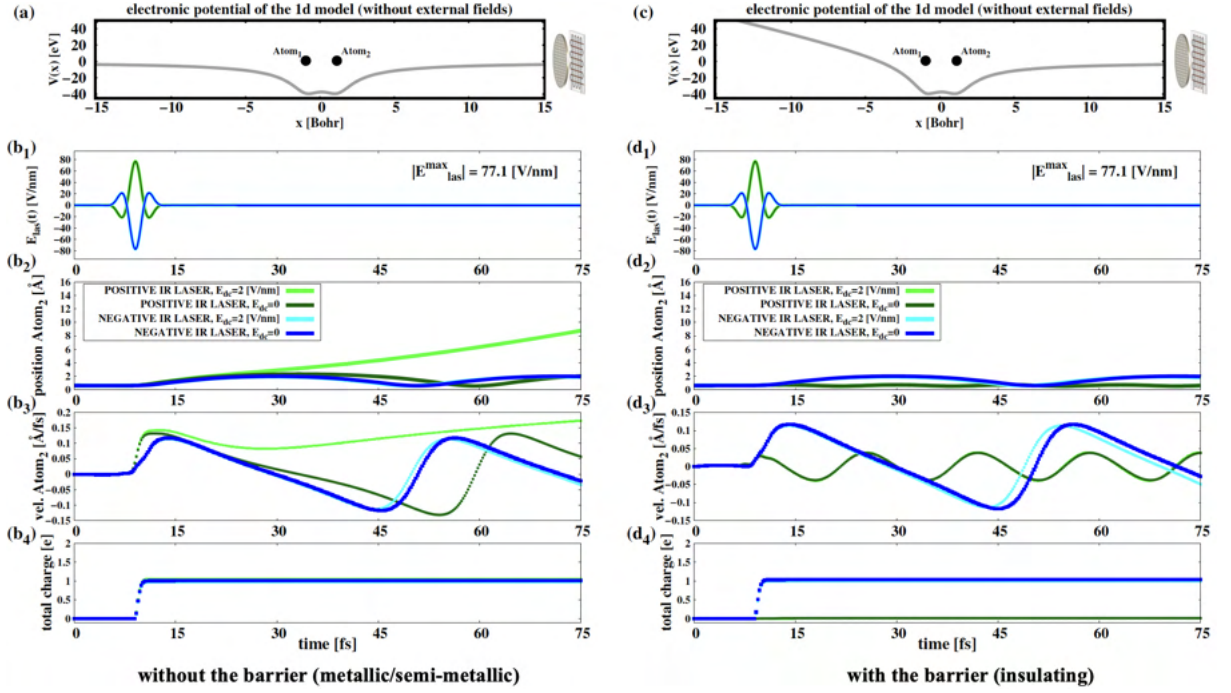


Figure B.3: Comparison of the results obtained with the two potentials shown in panels (a) and (c) is reported for a fixed amplitude of a single-cycle IR pulse with frequency $\omega = 0.799898$ eV, corresponding to a wavelength of $1.55 \mu\text{m}$, in panels (b1) and (d1). The maximum amplitude is 0.15 atomic units. The light and dark green curves correspond to the positive pulse, while the cyan and blue lines represent the negative pulse. Panels (b2) and (d2) show the position of Atom2 as a function of time, while panels (b3) and (d3) display its velocity under the different conditions. Finally, panels (b4) and (d4) report the total charge inside the simulation box. On the left-hand side (metal-like case, panels (a) and (b1–b4)), Atom2 evaporates when a positive pulse is applied in conjunction with the static external field (light green line), as a result of the strong ionization shown in panel (b4). On the right-hand side (insulating-like case, panels (c) and (d1–d4)), evaporation does not occur. This behavior is consistent with the findings for THz pulses reported in Fig. 8 of the main text, although a larger pulse amplitude is required in the present case. Adapted from Ref. [184].

Case of Study: The YTHDF1-Ebselen Compound Interaction

This Chapter contains details and preliminary results of a study cited in Chapter 7, whose aim is to find clues about the interaction between the YTHDF1 protein and one of its inhibitors, Ebselen, an organoselenium compound. Ebselen is a particular molecule, as it contains the Selenium (Se) element, and is also classified as a covalent drug. Thus, a first introductory section will be dedicated to the description of this particular family of compounds, with a specific focus on the Ebselen and its inhibitory interaction with the YTHDF1 protein. A second section presents the scientific aim that I had to pursue in this very large collaboration, followed by a third and fourth section describing the simulation protocol employed and some preliminary results, respectively. More details about the cited experimental procedures in this Chapter can be found in Refs. [231, 232], that also represent the fundamental basis of this study.

C.1 Introduction

C.1.1 Covalent Drugs

Covalent drugs form a covalent bond with their biological target. Although historically considered risky because of potential off-target reactivity and chemically reactive metabolites, they have become an important class of therapeutics, with successful examples including aspirin, penicillins and several kinase inhibitors. When properly designed, covalent inhibitors can achieve high potency, prolonged target inhibition and activity against targets that are difficult to modulate with conventional reversible ligands [233].

The mechanism of action of a covalent inhibitor generally involves two sequential steps, as follows:



First, the ligand binds reversibly to the target protein, forming a non-covalent complex characterized by the equilibrium constant $k_i = \frac{k_{\text{on}}}{k_{\text{off}}}$. This recognition step correctly positions the electrophilic warhead (a reactive electrophilic functional group) close to a nucleophilic amino

acid residue and largely determines target selectivity. Subsequently, the covalent bond is formed with rate constant k_2 , yielding either a reversible or irreversible covalent complex depending on the chemistry of the warhead. For irreversible inhibitors, the reverse rate constant (k_{-2}) is effectively zero, resulting in permanent target inactivation until new protein is synthesized.

Because inhibition depends on both binding affinity and covalent bond formation, conventional IC_{50} values alone are often insufficient to characterize covalent inhibitors. Their apparent potency is influenced by both k_i and k_2 , as well as by additional factors such as the lifetime of the covalent complex and the target protein turnover rate.

Selectivity primarily arises from the non-covalent recognition step rather than from highly reactive electrophiles. Once the ligand is correctly oriented within the binding pocket, the proximity between the electrophilic warhead and the target nucleophile greatly accelerates the reaction. Consequently, moderately reactive warheads are generally preferred over highly reactive electrophiles, which tend to react indiscriminately with off-target proteins and are frequently classified as PAINS [234]. Rational design therefore aims to combine a highly selective binding scaffold with an appropriately reactive warhead.

A wide variety of electrophilic warheads have been developed for targeted covalent inhibitors. Among these, α, β -unsaturated amides (Michael acceptors) are the most widely used because they efficiently react with cysteine residues while maintaining relatively low intrinsic reactivity. Other warheads have been designed to target cysteine, lysine, tyrosine, serine, threonine and additional nucleophilic amino acids, considerably expanding the applicability of covalent inhibition strategies [233, 235].

The irreversible nature of many covalent inhibitors provides sustained pharmacological effects, since recovery of target activity depends on *de novo* protein synthesis rather than drug clearance. Furthermore, covalent inhibitors can retain efficacy against certain resistance mutations, which often reduce the binding affinity of reversible inhibitors without completely preventing covalent bond formation.

C.1.2 The Ebselen Compound

Selenium is an essential trace element involved in redox homeostasis through several selenoproteins. Owing to its higher nucleophilicity, polarizability and redox activity compared to sulfur, organoselenium compounds have attracted considerable interest in medicinal chemistry [236, 237].

Ebselen (2-phenylbenzisoselenazol-3(2H)-one) is one of the best-known organoselenium compounds. It exhibits low toxicity and broad pharmacological activity, including antioxidant, antibacterial, antiviral and anticancer effects [238, 239]. Its antioxidant properties arise mainly from its glutathione peroxidase-like activity and modulation of the thioredoxin system.

Ebselen acts as a reversible covalent inhibitor by reaction of its electrophilic selenium atom with cysteine residues, forming a Se–S bond that can be cleaved under reducing conditions. This mechanism underlies its activity against several biological targets, including the SARS-CoV-2 main protease (M^{Pro}) [240] (see Fig. C.1). The favorable biological profile of ebselen has stimulated the synthesis of numerous selenium- and sulfur-based analogues with improved potency and selectivity.

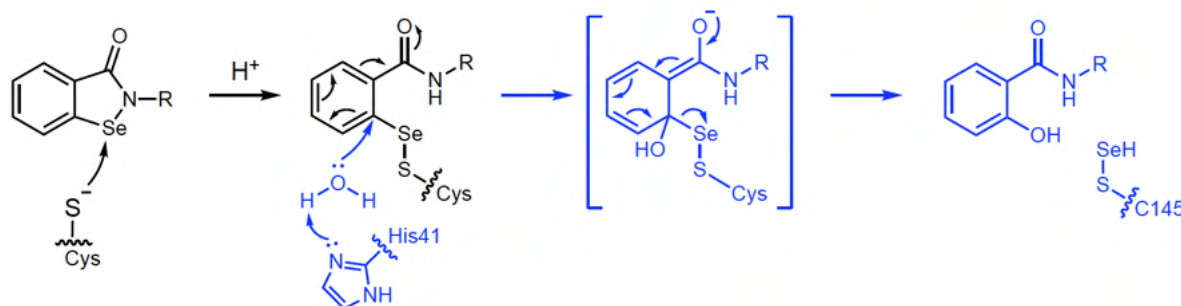


Figure C.1: The mechanism of the Se-S covalent bond formation by reaction with a cysteine residue on the electrophilic selenium atom in the SARS-CoV-2 main protease (M^{Pro}). In blue the peculiar histidine-catalyzed S_NAr-like mechanism, leading to the selenylation of Cys145. Adapted from Ref. [232].

C.1.3 Ebselen as a covalent inhibitor of YTHDF proteins

Researchers from the Bio-organic Chemistry Laboratory and the Laboratory of Protein Crystallography and Structure-Based Drug Design (University of Trento) reported in a previous study [231] that ebselen covalently binds a conserved cysteine residue located adjacent to the m⁶A recognition pocket of YTHDF1–3 proteins.

Specifically, ebselen was identified as a small-molecule inhibitor of YTHDF proteins through a tryptophan fluorescence quenching screening assay, exploiting the intrinsic fluorescence of the aromatic residues forming the m⁶A recognition pocket. After a screening of a library of 2560 bioactive molecules, this compound showed interference with the YTH domains of YTHDF1 and YTHDF2, while no interaction was observed with YTHDC1. This selectivity was attributed to the presence of a conserved cysteine residue (C412 in YTHDF1) located near the m⁶A-binding pocket (see Fig. C.2), which is replaced by serine in YTHDC1. Further biochemical and biophysical analyses confirmed that ebselen inhibits m⁶A-RNA recognition through a covalent mechanism involving its selenium reactive group. More details regarding all the employed experimental procedures can be found in Ref. [231].

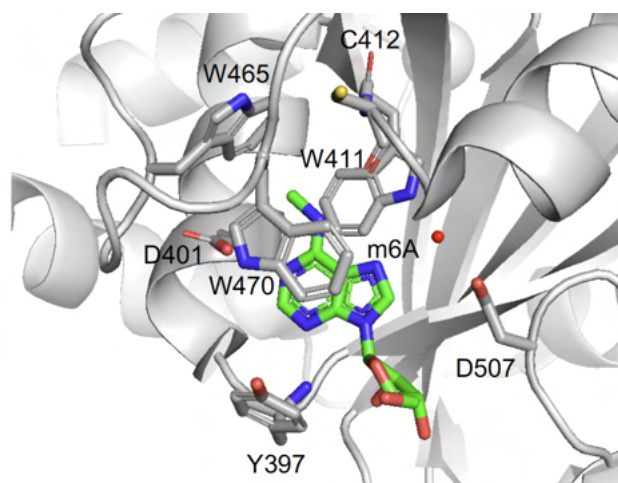


Figure C.2: Crystal structure of the YTH domain of the YTHDF1 protein (PDB code: 4RCJ). The m⁶A-modified RNA is contained in the PDB file and shown in green, with only the N⁶-methyladenosine residue highlighted (the resolved 3D profile contains three more nucleotides). The cysteine residue C412, located near the m⁶A recognition domain, is also shown. Adapted from Ref. [232].

The structural basis of ebselen activity was elucidated by X-ray crystallography, performed on the YTHDF1 YTH domain in complex with the ligand. The crystal structure confirmed the formation of a covalent bond between the selenium atom and C412, positioning the compound close to the m⁶A recognition site, mainly composed of three tryptophan residues which define a narrow hydrophobic pocket for the modified nucleotide. Ebselen adopts two slightly different binding conformations in the crystal structure (PDB code: 7PCU [241]) as seen in Fig. C.3, while maintaining interactions with residues surrounding the hydrophobic recognition groove. The N-phenylformamide moiety establishes contacts with D401, R404, S405, I410 and W411, whereas the selenobenzene group extends towards the region containing K471, G472 and K473.

C.1.4 The Relevance of the $\beta 4$ – $\beta 5$ Loop of YTHDF1 for the Inhibitory Mechanism

The inhibitory mechanism of ebselen is mainly related to the disruption of the conformational rearrangement of the $\beta 4$ – $\beta 5$ loop required for m⁶A recognition and stabilization. In the m⁶A-bound YTHDF1 structure, this loop adopts a defined conformation that positions W465 together with W411 and W470, forming the aromatic cage responsible for methylated adenine recognition, as seen in the 4RCJ PDB file (see Fig. C.2). Covalent attachment of ebselen to C412 prevents this rearrangement, resulting in loop destabilization and disorder. Fig C.4 displays a

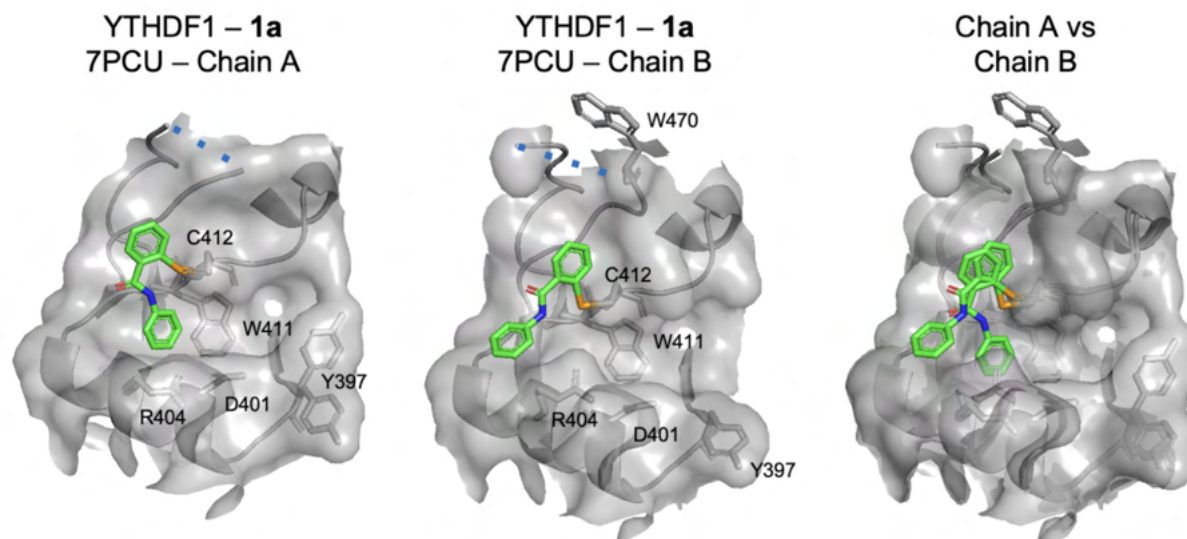


Figure C.3: Co-crystal structures of Ebselen (green) covalently bound to C412 of YTHDF1 domain (gray) (PDB code: 7PCU). The blue dotted lines highlight the disordered loop conformation (for this reason, not present in the resolved structure) that disrupts the m⁶A binding domain. A slightly different orientation can be noted in chain A (left) and chain B (middle). On the right, the overlap of the two chains emphasizes the different poses assumed by Ebselen. Adapted from Ref. [232].

structural comparison between the m⁶A-bound and ebselen-bound structures, revealing steric incompatibilities between the compound and residues S461, A462, V464 and W465, preventing the formation of the m⁶A-binding competent conformation. Moreover, ebselen binding induces the displacement of W465 and W470, generating a new pocket architecture compared with the native m⁶A-recognition state. This rearranged conformation provides a structural framework for the development of new YTHDF inhibitors exploiting the ebselen-induced binding site [231]. Based on these findings, additional selenium- and sulfur-containing analogues were and are currently being investigated to further optimize inhibition of m⁶A-RNA recognition by YTHDF proteins [242].

Interestingly, there is also another structure of YTHDF1 that was recently resolved. Compared with the previously reported YTHDF1 structures, PDB 7YZ8 [243] provides an alternative conformational state of the β 4– β 5 loop that differs from both the active m⁶A-bound (4RCJ) and the ebselen-bound (7PCU) structures.

In the canonical m⁶A-bound structure (PDB: 4RCJ), the β 4– β 5 loop adopts a closed conformation that positions W465 together with W411 and W470 to form the aromatic cage required for m⁶A recognition. In contrast, the β 4– β 5 loop in PDB: 7YZ8 remains well ordered but undergoes a pronounced conformational rearrangement, resulting in different orientations of W465

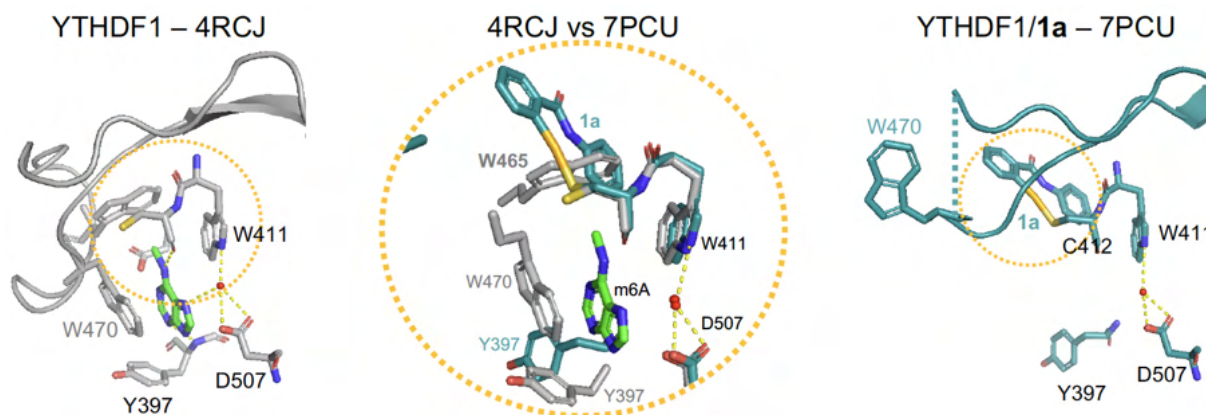


Figure C.4: Comparison of the $\beta 4$ – $\beta 5$ loop conformations in YTHDF1. Left: structured loop and m^6A recognition domain in the binding-competent conformation. Right: disordered loop induced by covalent binding of Ebselen (compound 1a in the image) to C412. The superposition highlights the displacement of W465 caused by the aromatic moiety of Ebselen and the rearrangement of W470. The positions of W411 and D507 remain largely conserved between the two structures. Adapted from Ref. [232].

and W470 and the establishment of an alternative network of aromatic interactions compared with the canonical m^6A -bound state. Consequently, the aromatic cage is not assembled in a binding-competent conformation, giving rise to a distinct binding pocket. Moreover, this rearrangement increases the solvent accessibility of the nearby C412 residue, potentially facilitating the binding of Ebselen.

Unlike the ebselen-bound structure (PDB: 7PCU), where covalent attachment of ebselen to C412 results in a largely disordered $\beta 4$ – $\beta 5$ loop and disruption of the aromatic cage, 7YZ8 demonstrates that the loop can adopt a stable alternative conformation in the absence of ligand binding.

To sum up, Fig. C.5 reports all the key details of the three YTHDF1 PDB files. These structures highlight the intrinsic conformational plasticity of the $\beta 4$ – $\beta 5$ loop and identify this region as a promising target for structure-based inhibitor design [244, 241, 245].

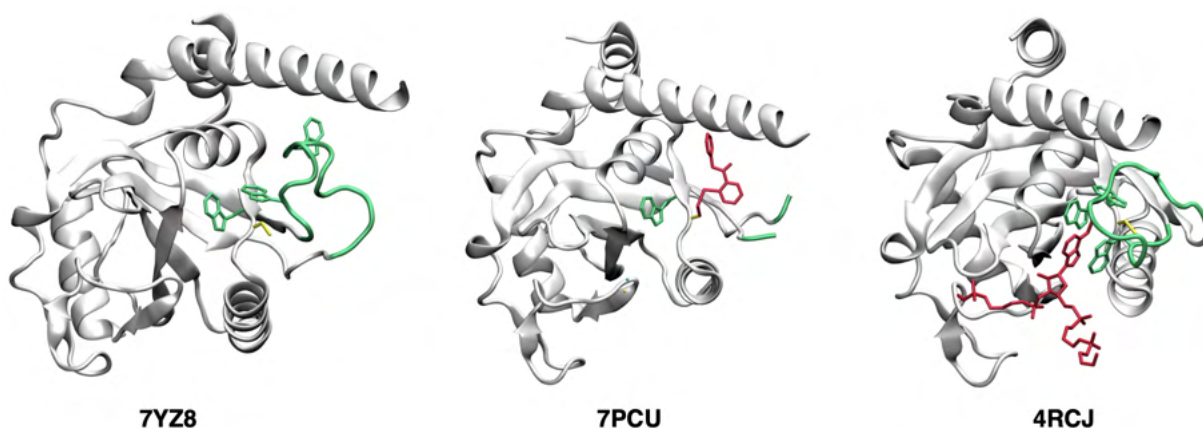


Figure C.5: Comparison of the three PDB files representing three different states of the $\beta 4$ - $\beta 5$ loop. Left: alternative conformational state of the $\beta 4$ - $\beta 5$ loop without presence of endogenous ligands or inhibitors. C412 is more accessible. Middle: unresolved (thus, disordered) loop induced by covalent binding of Ebselen to C412 (7PCU). Right: structured loop and m⁶A recognition domain in the binding-competent conformation (4RCJ), that completely covers C412. Color code: ebselen and the resolved piece of mRNA containing m⁶A (only the sidechain of the methylated Adenine is shown) are reported in red; the three Tryptophans constituting the m⁶A binding pocket (W411, W466 and W470) are displayed in green, together with the $\beta 4$ - $\beta 5$ loop; C412 is reported in yellow.

C.2 Aim of The Study

Up to here, available structural data indicate that the $\beta 4$ - $\beta 5$ loop plays a central role in regulating the accessibility of the YTHDF1 m⁶A-binding pocket. In the canonical m⁶A-bound structure (PDB: 4RCJ), the loop adopts a closed conformation that assembles the aromatic cage and buries C412 beneath the recognition site. Conversely, in the ebselen-bound structure (PDB: 7PCU), covalent attachment of the ligand to C412 is associated with a non-resolved $\beta 4$ - $\beta 5$ loop and disruption of the aromatic cage. Stating that the loop is completely disordered would be inaccurate without a structural proof, but at least it is reasonable to say that it is in an **open** form, so that C412 becomes accessible to ebselen. More recently, the alternative crystal structure 7YZ8 revealed that the loop can also populate a stable ordered conformation distinct from both 4RCJ and 7PCU, exposing a different pocket architecture and suggesting that the YTH domain samples multiple conformational states separated by significant free-energy barriers rather than behaving as a simple open/closed switch [241, 246, 247].

These observations raise a fundamental mechanistic question: *how does ebselen reach C412 to initiate covalent bond formation?* If the active m⁶A-bound conformation (4RCJ) is maintained also without the presence of the mRNA or also if the loop stays in the 7YZ8 form, the cysteine appears poorly accessible to solvent and ligand. Therefore, before investigating the chemical

reaction leading to covalent bond formation, it is essential to characterize the preceding non-covalent recognition event, that should bring the compound close to C412. As described at the beginning of this Chapter, in a typical covalent inhibition mechanism, the first affinity (k_i) is determined by the initial reversible binding step, whereas the subsequent covalent reaction is described by the inactivation rate constant (k_2), assuming k_{-2} null. The present study therefore focuses on the first step of this process, namely the structural pathway through which ebselen can access C412: in other words, K_i should be estimated somehow.

To address this question, plain molecular dynamics and also localized volume-based well-tempered metadynamics simulations were performed starting from the three available structural states (PDB: 4RCJ, 7YZ8 and 7PCU), representing three completely different conformations of the loop, that will be named **closed** (4RCJ), **semi-open** (7YZ8), and **open** –or disordered– (7PCU), respectively. By selectively enhancing the conformational sampling of Ebselen and the $\beta 4$ – $\beta 5$ loop, these simulations aim to characterize and quantify the accessibility of C412, and identify which structural state provides the most favorable pathway for the initial recognition of ebselen prior to covalent bond formation.

C.3 Simulation Protocol

C.3.1 Crystal Structures

The crystal structures of the three systems investigated in this study (4RCJ, 7PCU, and 7YZ8) were obtained from the RCSB Protein Data Bank [124]. For 7PCU, missing residues of the loop were reconstructed employing the Modeller interface [130] in UCSF Chimera [131]. Moreover, the covalent-bound Ebselen molecule was removed. The purpose of this procedure is to have a conformation of the loop as open as possible, so that the C412 is in principle completely accessible (this is also why it was not employed a more advanced software i.e. Machine Learning-based to reconstruct this loop). Fig. C.6 shows the differences between the original and the reconstructed 7PCU PDB file.

C.3.2 Force Fields Selection and Compatibility

The Amber ff14SB [139] force field was used to model the biomolecules, TIP3P [140] was employed for water molecules, and GAFF was adopted for Ebselen. Sodium and chloride ions were described using the Joung–Cheatham ion parameters [141].

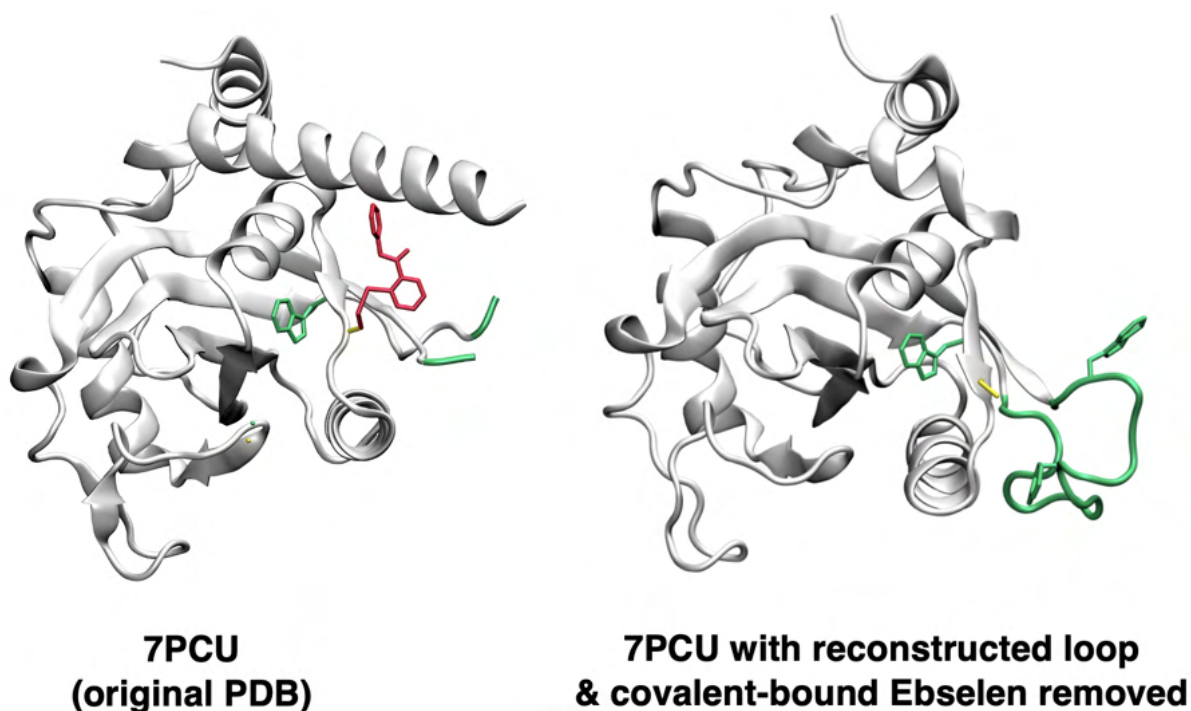


Figure C.6: Comparison of the two 7PCU PDB files representing the original PDB file (left) and the reconstructed $\beta 4$ – $\beta 5$ loop without the Ebselen molecule covalently bound (right). Color code: the three Tryptophans constituting the m6A binding pocket (W411, W466 and W470) are displayed in green, together with the $\beta 4$ – $\beta 5$ loop; C412 is reported in yellow.

C.3.3 Plain MD setup

The protonation states of all biomolecules were evaluated and verified using the H++ online server [142]. Classical all-atom molecular dynamics simulations were performed with GROMACS 2021.5 [143] in explicit water containing 150 mM NaCl. All simulated systems underwent the following equilibration protocol: (i) 50,000 steps of energy minimization using the steepest descent algorithm; (ii) a 1 ns NVT equilibration at $T = 298$ K using a Berendsen thermostat [144] with a coupling constant of $\tau_T = 0.1$ ps; and (iii) a subsequent 2 ns NPT equilibration at $P = 1$ bar using a Berendsen barostat [144], with $\tau_P = 2.0$ ps and a water compressibility of $4.5 \cdot 10^{-5}$ bar $^{-1}$. During these equilibration stages, all heavy atoms of the solute were restrained with a force constant of 1000 kJ/(mol $\cdot$ nm), and a 1 fs integration time step (t_{step}) was employed.

Following equilibration, each system was subjected to an additional 5 ns unrestrained NPT equilibration, after which three independent production simulations of 1 μ s each were performed, yielding a cumulative simulation time of 3.0 μ s per system. Production runs were carried out in the NPT ensemble using the stochastic velocity-rescale thermostat [109] ($\tau_T = 0.1$ ps, $T = 298$

K) together with the Parrinello–Rahman barostat [145] ($\tau_P = 2.0$ ps, water compressibility = $4.5 \cdot 10^{-5}$ bar $^{-1}$, $P = 1$ bar). An integration time step of 2 fs was used throughout the production phase.

Electrostatic interactions were calculated using the particle mesh Ewald (PME) method [146], with a 1.0 nm cutoff applied to both short-range electrostatic and van der Waals interactions. Covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm [147], while long-range dispersion corrections were included in both the energy and pressure calculations [148].

C.3.4 Localized Volume-Based Metadynamics

To investigate the structural determinants governing the accessibility of C412 to ebselen, localized volume-based well-tempered metadynamics (LV-MetaD) was employed using the PLUMED plugin interfaced with GROMACS. This enhanced sampling approach, originally proposed by Capelli *et al.*, is specifically designed to investigate ligand recognition processes by restricting the exploration of the ligand to a predefined volume surrounding the binding site, thereby avoiding the inefficient sampling of bulk solvent while preserving an unbiased description of ligand binding and unbinding events [248, 249, 250].

Unlike conventional metadynamics, where the ligand is free to diffuse throughout the entire simulation box, LV-MetaD confines the sampling within a geometrically defined region encompassing the protein binding site. This strategy substantially accelerates the exploration of the configurational space relevant to the binding process while maintaining the possibility of observing multiple binding and unbinding events. Since the present work aims to characterize the structural pathway through which ebselen reaches the reactive residue C412, rather than the covalent reaction itself, LV-MetaD represents an appropriate approach to investigate the initial reversible recognition step preceding covalent bond formation. The simulation protocol is the same of that of plain MD, with the difference that all the production runs were run coupled with this enhanced method algorithm.

Collective variables

The position of ebselen relative to the protein was described using a spherical coordinate system centered on a reference region of the protein. A group of structurally rigid C_α atoms ($\text{RMSF} < 0.20$ nm) was used to define the reference frame, while the ligand center of mass was employed to describe ligand motion. After optimal fitting of the protein onto the reference structure, the ligand coordinates were transformed into the three collective variables:

$$\rho = \sqrt{x^2 + y^2 + z^2} \quad (\text{C.2})$$

$$\theta = \tan^{-1} \left(\frac{y}{x} \right) \quad (\text{C.3})$$

$$\tau = \sqrt{\rho - z}. \quad (\text{C.4})$$

The variables ρ , θ and τ respectively describe the radial distance, the azimuthal angle and the vertical position of the ligand relative to the protein reference frame. Metadynamics bias was deposited simultaneously along these three coordinates, allowing the ligand to freely explore the accessible volume surrounding the binding pocket.

Paraboloidal restraining volume

The localized volume was generated through restraining potentials acting on ρ and τ . While the upper wall on ρ limits the maximum radial distance explored by the ligand, the definition of τ naturally generates a paraboloidal boundary. From the definition

$$\tau = \sqrt{\rho - z}, \quad (\text{C.5})$$

it follows that

$$\rho = \tau^2 + z. \quad (\text{C.6})$$

Considering the definition of the spherical radius,

$$\rho^2 = x^2 + y^2 + z^2, \quad (\text{C.7})$$

substitution yields

$$x^2 + y^2 = \tau^4 + 2\tau^2 z, \quad (\text{C.8})$$

which corresponds to the equation of a paraboloid of revolution. Therefore, imposing an upper wall on τ confines ligand diffusion inside a paraboloidal volume aligned with the binding pocket entrance, while avoiding the sharp discontinuities introduced by purely spherical restraints.

In addition, an upper wall was applied to the protein RMSD to preserve the native fold during the enhanced sampling simulations and to ensure that the observed ligand motions resulted from local conformational fluctuations of the binding site rather than global structural distortions.

The restraining potentials were therefore defined as

- radial restraint: $\rho \leq 3.5$ nm;
- paraboloidal restraint: $\tau \leq \tau_{\text{max}}$;
- protein stability restraint: $\text{RMSD} \leq 0.10$ nm.

The harmonic force constants were set to $22000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ for the radial restraint and $20000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ for both the paraboloidal and RMSD restraints, preventing ligand escape from the localized sampling region while minimally perturbing the conformational dynamics inside the explored volume.

Well-tempered Metadynamics

Bias deposition was performed using the well-tempered metadynamics formalism [249]. Gaussian potentials with an initial height of 1.2 kJ mol^{-1} were deposited every 500 simulation steps (1 ps) along the three collective variables ρ , τ and θ . Gaussian widths were set to 0.10 nm for ρ , $0.04 \text{ nm}^{1/2}$ for τ , and $\pi/8$ rad for θ . A bias factor of 10 was employed at 310 K, ensuring gradual filling of the free-energy landscape while preserving convergence towards the underlying free-energy surface.

Simulation design of LV-MetaD MD Simulations

Localized volume-based metadynamics simulations were carried out starting from the three experimentally determined conformational states of the YTHDF1 YTH domain:

- PDB 4RCJ (m⁶A-bound conformation);
- PDB 7YZ8 (alternative ordered loop conformation);
- PDB 7PCU (ebselen-bound conformation).

To investigate the effect of the size of the localized sampling region on ligand accessibility, three different values of the paraboloidal restraint were considered, $\tau_{\text{max}} = 0.6, 0.8, 1.0$, corresponding to progressively larger accessible sampling volumes surrounding the binding pocket. For each combination of starting structure and τ_{max} , three independent replicas were performed using different initial velocity distributions.

Each simulation consisted of a $1 \mu\text{s}$ production run, resulting in a total of 27 independent localized metadynamics simulations, $3 \text{ structures} \times 3 \tau_{\text{max}} \times 3 \text{ replicas} = 27$, corresponding to an aggregate simulation time of $27 \mu\text{s}$. The overall objective of this simulation campaign was to correlate the three experimentally observed conformational states of the $\beta 4$ – $\beta 5$ loop with the accessibility of C412 and the ability of ebselen to enter the YTHDF1 binding pocket, thereby identifying the structural state that most likely represents the productive recognition pathway preceding covalent bond formation.

C.4 Preliminary Results and Conclusions

***Note:** At present, the associated computational simulations and data analyses are still ongoing, and therefore a comprehensive presentation of the scientific findings would be premature. It is expected that the complete results will become available in the near future, allowing their dissemination through future scientific publications.*

References

- [1] MS Smyth and JHJ Martin. x ray crystallography. *Molecular Pathology*, 53(1):8, 2000.
- [2] Eran Eyal, Sergey Gerzon, Vladimir Potapov, Marvin Edelman, and Vladimir Sobolev. The limit of accuracy of protein modeling: influence of crystal packing on protein structure. *Journal of molecular biology*, 351(2):431–442, 2005.
- [3] Naomi E Chayen. Turning protein crystallisation from an art into a science. *Current opinion in structural biology*, 14(5):577–583, 2004.
- [4] Martin Caffrey and Vadim Cherezov. Crystallizing membrane proteins using lipidic mesophases. *Nature protocols*, 4(5):706–731, 2009.
- [5] Umesh B Gawas, Vinod K Mandrekar, and Mahesh S Majik. Structural analysis of proteins using x-ray diffraction technique. In *Advances in biological science research*, pages 69–84. Elsevier, 2019.
- [6] James A. Geraets, Karunakar R. Pothula, and Gunnar F. Schröde. Integrating cryo-em and nmr data. *Current Opinion in Structural Biology*, 61:173–181, 2020.
- [7] Philippe Cuniasse, Paulo Tavares, Elena V. Orlova, and Sophie Zinn-Justin. Structures of biomolecular complexes by combination of nmr and cryoem methods. *Current Opinion in Structural Biology*, 43:104–113, 2017.
- [8] Xiao chen Bai, Greg McMullan, and Sjors H. W. Scheres. How cryo-em is revolutionizing structural biology. *Trends in Biochemical Sciences*, 40:49–57, 2015.
- [9] Phoebe L. Stewart. Cryo-electron microscopy and cryo-electron tomography of nanoparticles. *WIREs Nanomedicine and Nanobiotechnology*, 9:e1417, 2017.
- [10] Rafael Fernandez-Leiro and Sjors H. W. Scheres. Unravelling biological macromolecules with cryo-electron microscopy. *Nature Reviews Molecular Cell Biology*, 2016.
- [11] BioRender. <https://www.biorender.com>.
- [12] Wen Jiang, Matthew L. Baker, Joanita Jakana, Peter R. Weigele, Jonathan King, and Wah Chiu. Backbone structure of the infectious ϵ 15 virus capsid revealed by electron cryomicroscopy. *Nature*, 451:1130–1134, 2008.

- [13] X. Yu, L. Jin, and Z. H. Zhou. 3.88 Å structure of cytoplasmic polyhedrosis virus by cryo-electron microscopy. *Nature*, 453:415–419, 2008.
- [14] Phoebe L. Stewart. Cryo-electron microscopy and cryo-electron tomography of nanoparticles. *WIREs Nanomedicine and Nanobiotechnology*, 9:e1417, 2017.
- [15] Peter Ercius, Jihan Zhou, Yongsoo Yang, Yao Yang, Dennis Kim, Andrew Yuan, Xuezheng Tian, Colin Ophus, S. Zhu, Andreas Schmid, Michael Nathanson, Hendrik Heinz, Qi An, Hao Zeng, and Jianwei Miao. Imaging nucleation, growth and disorder at the single-atom level by atomic electron tomography (aet). *Microscopy and Microanalysis*, 26:1848–1850, 2020.
- [16] Yong Zi Tan et al. Addressing preferred specimen orientation in single-particle cryo-em through tilting. *Nature Methods*, 14(8):793–796, 2017.
- [17] M. A. Aronova and R. D. Leapman. Development of electron energy-loss spectroscopy in the biological sciences. *MRS Bulletin*, 37(1):53–62, 2012.
- [18] R. F. Egerton. *Electron Energy-Loss Spectroscopy in the Electron Microscope*. Springer, New York, 3 edition, 2011.
- [19] R. F. Egerton. Radiation damage to organic and inorganic specimens in the tem. *Micron*, 119:72–87, 2019.
- [20] Baptiste Gault et al. Atom probe tomography. *Nature Reviews Methods Primers*, 1:51, 2021.
- [21] M. K. Miller and R. G. Forbes. *Atom-Probe Tomography: The Local Electrode Atom Probe*. Springer Science+Business Media, New York, 2014.
- [22] Baptiste Gault et al. *Atom Probe Microscopy*, volume 160 of *Springer Series in Materials Science*. Springer, New York; Heidelberg; Dordrecht; London, 2012.
- [23] Lefebvre Williams, François Vurpillot, and Xavier Sauvage, editors. *Atom Probe Tomography: Put Theory Into Practice*. Academic Press (Elsevier), 2016.
- [24] Y.-S. Chen et al. Atom probe tomography for the observation of hydrogen in materials: A review. *Microscopy and Microanalysis*, 29(1):1–28, 2023.
- [25] T. F. Kelly and M. K. Miller. On the many advantages of local-electrode atom probes. *Ultramicroscopy*, 62(1-2):29–42, 1996.
- [26] D. J. Larson et al. *Local Electrode Atom Probe Tomography*. Springer, New York, NY, USA, 2013.
- [27] Robert Gomer. *Field Emission and Field Ionization*, volume 34. Springer, 1961.
- [28] Toshio Sakurai and Erwin W. Muller. Field calibration using the energy distribution of a free-space field ionization. *Journal of Applied Physics*, 48(6):2618–2625, 1977.

- [29] Francois Vurpillot et al. Pragmatic reconstruction methods in atom probe tomography. *Ultramicroscopy*, 111(8):1286–1294, 2011.
- [30] Francois Vurpillot. Field ion emission mechanisms. In *Atom Probe Tomography*, pages 17–72. Elsevier, 2016.
- [31] Shyeh Tjing Loi et al. Electrostatic simulations of a local electrode atom probe: The dependence of tomographic reconstruction parameters on specimen and microscope geometry. *Ultramicroscopy*, 132:107–113, 2013.
- [32] Cristelle Pareige et al. Time-of-flight mass spectrometry and composition measurements. In *Atom Probe Tomography*, pages 123–154. Elsevier, 2016.
- [33] Michael K. Miller. *Atom Probe Tomography: Analysis at the Atomic Level*. Springer Science & Business Media, 2012.
- [34] Ty J. Prosa and David J. Larson. Modern focused-ion-beam-based site-specific specimen preparation for atom probe tomography. *Microscopy and Microanalysis*, 23(2):194–209, 2017.
- [35] Gérald Da Costa. Atom probe tomography: Detector issues and technology. In *Atom Probe Tomography*, pages 155–181. Elsevier, 2016.
- [36] P. Bas et al. Applied surface science. *Applied Surface Science*, 87-88:298–304, 1995.
- [37] E. W. Müller. Field ion microscopy. *Physical Review*, 102(3):618–624, 1956.
- [38] R. Gomer. Field emission and field ionization. *Journal of Chemical Physics*, 31:341, 1959.
- [39] R. Gomer and L. W. Swanson. Field evaporation of metals. *Journal of Chemical Physics*, 38:1613, 1963.
- [40] R. Haydock and D. R. Kingham. Post-ionization of field-evaporated ions. *Physical Review Letters*, 44:1520–1523, 1980.
- [41] D. R. Kingham. The post-ionization of field-evaporated ions: A theoretical explanation of multiple charge states. *Surface Science*, 116(2):273–301, 1982.
- [42] A. Menand and D. Blavette. In: *Le Journal de Physique Colloques*, 47(C7):17–20, 1986.
- [43] A. Menand and D. R. Kingham. Isotopic variations in field evaporation charge-state of boron ions. *J. Phys. D: Appl. Phys.*, 17(1):203–208, 1984.
- [44] G. L. Kellogg. Determining the field emitter temperature during laser irradiation in the pulsed laser atom probe. *Journal of Applied Physics*, 52:5320–5328, 1981.
- [45] R. K. Biswas and R. G. Forbes. Field evaporation: theory and analysis. *Journal of Physics D: Applied Physics*, 15(7):1323–1338, 1982.

- [46] K. Hirose and M. Tsukada. First-principles theory of field evaporation. *Physical Review Letters*, 73(1):150, 1994.
- [47] H. J. Kreuzer, L. C. Wang, and N. D. Lang. Quantum theory of field evaporation. *Physical Review B*, 45(20):12050–12055, 1992.
- [48] N. D. Lang. Field evaporation: quantum-mechanical aspects. *Solid State Communications*, 84(1–2):155–158, 1992.
- [49] N. D. Lang. Theory of field evaporation. *Physical Review B*, 45(23):13599, 1992.
- [50] N. D. Lang. Microscopic theory of field evaporation. *Physical Review B*, 49(3):2067, 1994.
- [51] Y. Suchorski et al. Surface processes in strong electric fields. *Progress in Surface Science*, 48(1–4):121–134, 1995.
- [52] Y. Suchorski et al. Field evaporation and related surface phenomena. *Progress in Surface Science*, 53(2–4):135–153, 1996.
- [53] C. G. Sanchez, A. Y. Lozovoi, and A. Alavi. Density functional theory of field evaporation. *Molecular Physics*, 102(9–10):1045–1055, 2004.
- [54] T. T. Tsong and C. S. Chang. Field evaporation and post-ionization: a critical reappraisal. *Japanese Journal of Applied Physics*, 34(6B):3309–3318, 1995.
- [55] G. L. Kellogg. Measurement of activation energies for field evaporation of tungsten ions as a function of electric field. *Physical Review B*, 29(8):4304–4312, 1984.
- [56] T. T. Tsong et al. Photon-assisted field ionization. *The Journal of Chemical Physics*, 65(6):2469–2470, 1976.
- [57] B. Viswanathan et al. Laser-induced field ionization processes. *The Journal of Chemical Physics*, 70:2582, 1979.
- [58] S. Nishigaki, W. Drachsel, and J. H. Block. Photon-induced field ionization mass spectrometry of ethylene on silver. *Surface Science*, 87(2):389–409, 1979.
- [59] G. L. Kellogg and T. T. Tsong. Laser-pulsed atom probe microscopy. *Ultramicroscopy*, 5(2):259–260, 1980.
- [60] W. Drachsel, S. Nishigaki, and J. H. Block. Laser-assisted field ionization mechanisms. *International Journal of Mass Spectrometry and Ion Physics*, 32(4):333–343, 1980.
- [61] T. T. Tsong, S. B. McLane, and T. Kinkus. Pulsed-laser atom probe instrumentation. *Review of Scientific Instruments*, 53(9):1442–1448, 1982.
- [62] P. Stender et al. Sub-nanosecond laser pulsing in atom probe tomography. *Ultramicroscopy*, 107(9):726–733, 2007.

- [63] J. H. Bunton et al. Picosecond laser-assisted atom probe tomography. *Microscopy and Microanalysis*, 13:418–427, 2007.
- [64] A. Cerezo, G. D. W. Smith, and P. H. Clifton. Femtosecond laser-assisted atom probe tomography. *Applied Physics Letters*, 88(15):154103, 2006.
- [65] M. Bass et al. Optical rectification. *Physical Review Letters*, 9(11):446–448, 1962.
- [66] T. T. Tsong and T. J. Kinkus. Energy distributions of pulsed-laser field-desorbed gaseous ions and field-evaporated metal ions: a direct time-of-flight measurement. *Physical Review B*, 29(2):529, 1984.
- [67] T. T. Tsong. Pulsed-laser-stimulated field ion emission from metal and semiconductor surfaces: A time-of-flight study of the formation of atomic, molecular, and cluster ions. *Physical Review B*, 30(9):4946–4958, 1984.
- [68] J. G. Fujimoto et al. *Physical Review Letters*, 53:1837, 1984.
- [69] T. T. Tsong. Field penetration and band bending near semiconductor surfaces in high electric fields. *Surface Science*, 81(1):28–42, 1979.
- [70] Baishaki Mazumder et al. Reneutralization time of surface silicon ions on a field emitter. *New Journal of Physics*, 12(11):113029, 2010.
- [71] M. A. Fortes and B. Ralph. *Philosophical Magazine*, 18(154):787–805, 1968.
- [72] K. D. Rendulic and E. W. Müller. *Journal of Applied Physics*, 37(7):2593–2596, 1966.
- [73] M. Gilbert et al. *Ultramicroscopy*, 107(9):767–772, 2007.
- [74] M. J. G. Lee et al. Thermally enhanced field emission from a laser-illuminated tungsten tip: temperature rise of tip. *Journal of Applied Physics*, 51(9):4996–5006, 1980.
- [75] J. Hohlfield et al. Electron and lattice dynamics following optical excitation of metals. *Chemical Physics*, 251(1-3):237–258, 2000.
- [76] J. Houard et al. *Physical Review B*, 81(12):125411, 2010.
- [77] G. Sha, A. Cerezo, and G. D. W. Smith. Field evaporation behavior during irradiation with picosecond laser pulses. *Applied Physics Letters*, 92(4), 2008.
- [78] F. Vurpillot et al. *Journal of Physics D: Applied Physics*, 42(12):125502, 2009.
- [79] J. Houard et al. Three-dimensional thermal response of a metal subwavelength tip under femtosecond laser illumination. *Physical Review B*, 84(3):033405, 2011.
- [80] E. P. Silaeva et al. Do dielectric nanostructures turn metallic in high-electric dc fields? *Nano Letters*, 14(11):6066–6072, 2014.
- [81] Janusz Bogdanowicz et al. Laser-assisted atom probe tomography of semiconductors: The impact of the focused-ion beam specimen preparation. *Ultramicroscopy*, 188:19–23, 2018.

- [82] Gustav Sundell, Mats Hulander, Astrid Pihl, and Martin Andersson. Atom probe tomography for 3d structural and chemical analysis of individual proteins. *Small*, page 1900316, 2019.
- [83] K. Thompson, D. Lawrence, D. J. Larson, J. D. Olson, T. F. Kelly, and B. Gorman. In situ site-specific specimen preparation for atom probe tomography. *Ultramicroscopy*, 107:131–139, 2007.
- [84] M. Fragai, C. Luchinat, T. Martelli, E. Ravera, I. Sagi, I. Solomonov, and Y. Udi. Ss-nmr of biosilica-entrapped enzymes permits an easy assessment of preservation of native conformation in atomic detail. *Chemical Communications*, 50:421, 2014.
- [85] M. K. Miller, K. F. Russell, and G. B. Thompson. Strategies for fabricating atom probe specimens with a dual beam fib. *Ultramicroscopy*, 102:287, 2005.
- [86] L. E. Rayner, N. Kadkhodayi-Kholghi, R. K. Heenan, J. Gor, P. A. Dalby, and S. J. Perkins. The solution structure of rabbit igg accounts for its interactions with the fc receptor and complement c1q and its conformational stability. *Journal of Molecular Biology*, 425:506, 2013.
- [87] C. J. Brinker. Hydrolysis and condensation of silicates: effects on structure. *Journal of Non-Crystalline Solids*, 100:31–50, 1988.
- [88] C. A. Diebolder, F. J. Beurskens, R. N. de Jong, R. I. Koning, K. Strumane, M. A. Lindorfer, et al. Complement is activated by igg hexamers assembled at the cell surface. *Science*, 343:1260, 2014.
- [89] J. Belton, O. Deschaume, and C. C. Perry. An overview of the fundamentals of the chemistry of silica with relevance to biosilicification and technological advances. *FEBS Journal*, 279:1710, 2012.
- [90] E. O. Saphire, P. W. H. I. Parren, R. Pantophlet, M. B. Zwick, G. M. Morris, P. M. Rudd, et al. Crystal structure of a neutralizing human igg against hiv-1: a template for vaccine design. *Science*, 293:1155, 2001.
- [91] D. K. Schreiber, D. E. Perea, J. V. Ryan, J. E. Evans, and J. D. Vienna. A method for site-specific and cryogenic specimen fabrication of liquid/solid interfaces for atom probe tomography. *Ultramicroscopy*, 194:89, 2018.
- [92] R. Obert and B. C. Dave. Enzymatic conversion of carbon dioxide to methanol: enhanced methanol production in silica sol-gel matrices. *Journal of the American Chemical Society*, 121:12192, 1999.
- [93] A. Vella, J. Houard, L. Arnoldi, M. Tang, M. Boudant, A. Ayoub, A. Normandgerlad, G. Da Costa, and A. Hideur. High-resolution terahertz-driven atom probe tomography. *Science Advances*, 7:eabd7259, 2021.
- [94] T. F. Kelly. Kinetic-energy discrimination for atom probe tomography. *Microscopy and Microanalysis*, 17:1, 2011.

- [95] T. F. Kelly, M. K. Miller, K. Rajan, and S. P. Ringer. Atomic-scale tomography: A 2020 vision. *Microscopy and Microanalysis*, 19:652, 2013.
- [96] M. Karplus and G. A. Petsko. Molecular dynamics simulations in biology. *Nature*, 347:631–639, 1990.
- [97] Y. Xia, L. Tang, X. Lu, and S. Zhu. Theoretical insights into laser-assisted field evaporation of ionic compounds. *Journal of Applied Physics*, 136:134304, 2024.
- [98] Carsten A. Ullrich. *Time-Dependent Density-Functional Theory: Concepts and Applications*. Oxford Graduate Texts. Oxford University Press, Oxford, 2012.
- [99] Dominik Marx and Jürg Hutter. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press, Cambridge, 2009.
- [100] E. P. Silaeva, K. Uchida, Y. Suzuki, and K. Watanabe. Energetics and dynamics of laser-assisted field evaporation: Time-dependent density functional theory simulations. *Physical Review B*, 92:155401, 2015.
- [101] P. J. Feibelman. Surface-diffusion mechanism versus electric field: Pt/pt(001). *Physical Review B*, 64:125403, 2001.
- [102] C. G. Sanchez, A. Y. Lozovoi, and A. Alavi. Field-evaporation from first-principles. *Molecular Physics*, 102:1045–1057, 2004.
- [103] H. Tamura, M. Tsukada, K. P. McKenna, A. L. Shluger, T. Ohkubo, and K. Hono. Laser-assisted field evaporation from insulators triggered by photoinduced hole accumulation. *Physical Review B*, 86:195430, 2012.
- [104] Mark E. Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. Oxford University Press, New York, 2010.
- [105] Errol G. Lewars. *Computational Chemistry: Introduction to the Theory and Applications of Molecular and Quantum Mechanics*. Springer Science and Business Media B.V., 2011.
- [106] Luca Monticelli and D. Peter Tieleman. Force fields for classical molecular dynamics. In *Biomolecular Simulations*, volume 924 of *Methods in Molecular Biology*. Humana Press, Totowa, NJ, 2013.
- [107] Adri C. T. van Duin, Siddharth Dasgupta, François Lorant, and William A. III Goddard. Reaxff: A reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.
- [108] Thomas P. Senftle, Sungwook Hong, Md Mahbubul Islam, Sudhir B. Kylasa, Yuanxia Zheng, Yun Kyung Shin, Chad Junkermeier, Roman Engel-Herbert, Michael J. Janik, Hasan Metin Aktulga, Toon Verstraelen, Ananth Grama, and Adri C. T. van Duin. The reaxff reactive force-field: development, applications and future directions. *npj Computational Materials*, 2:15011, 2016.

- [109] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 01 2007.
- [110] Furio Ercolessi and James B Adams. Interatomic potentials from first-principles calculations: the force-matching method. *Europhysics Letters*, 26(8):583–588, 1994.
- [111] M Scott Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics*, 129(14):144108, 2008.
- [112] Siewert J Marrink, H Jelger Risselada, Sergei Yefimov, D Peter Tieleman, and Alex H de Vries. The martini force field: coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.
- [113] Paulo C T Souza, Riccardo Alessandri, Jonathan Barnoud, Sebastian Thallmair, Inês Faustino, Fabian Grünewald, Ioannis Patmanidis, Hamed Abdizadeh, Bart M H Bruininks, Tsjerk A Wassenaar, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature Methods*, 18(4):382–388, 2021.
- [114] Richard M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge University Press, Cambridge, United Kingdom, 2nd edition, 2020.
- [115] Giovanni Novi Inverardi, Francesco Carnovale, Lorenzo Petrolli, Simone Taioli, and Gianluca Lattanzi. Silica in silico: A molecular dynamics characterization of the early stages of protein embedding for atom probe tomography. *Biophysica*, 3(2):276–287, 2023.
- [116] Giovanni Novi Inverardi, Lorenzo Petrolli, Francesco Carnovale, Alessio Bartocci, Simone Taioli, and Gianluca Lattanzi. Adsorption of silica oligomers on biomolecules: Structural and dynamical insights for atom probe tomography via classic molecular dynamics simulations. *Computational and Structural Biotechnology Journal*, 2025.
- [117] P. F. Holt and J. E. L. Bowcott. The interaction of proteins with silicic acid. *Biochemical Journal*, 57(3):471–475, 1954.
- [118] S. G. Clark, P. F. Holt, and C. W. Went. The interaction of silicic acid with insulin, albumin and nylon monolayers. *Transactions of the Faraday Society*, 53:1500–1508, 1957.
- [119] S. G. Clark and P. F. Holt. The interaction of silicic acid with collagen and gelatin monolayers. *Transactions of the Faraday Society*, 53:1509–1515, 1957.
- [120] P. F. Holt. The polymerization of silicic acid and its subsequent interaction with proteins and other hydrogen-bonding agents. *Journal of Colloid Science*, 17(9):823–837, 1962.
- [121] V. Misra, Q. Rahman, and P. N. Viswanathan. Binding of silicic acid by proteins and its relation to toxicity of silicate dusts. *Journal of Applied Toxicology*, 3(3):135–138, 1983.
- [122] S. Parviainen, F. Djurabekova, S. P. Fitzgerald, A. Ruzibaev, and K. Nordlund. Atomistic simulations of field assisted evaporation in atom probe tomography. *Journal of Physics D: Applied Physics*, 49(4):045302, 2015.

- [123] S. J. Chen, X. Yao, C. Zheng, and W. H. Duan. Quantification of evaporation induced error in atom probe tomography using molecular dynamics simulation. *Ultramicroscopy*, 182:28–35, 2017.
- [124] Peter W. Rose, Andreas Prlić, A. Altunkaya, C. Bi, A. R. Bradley, C. H. Christie, et al. The rcsb protein data bank: integrative view of protein, gene and 3d structural information. *Nucleic Acids Research*, page gkw1000, 2016.
- [125] S. Vijay-Kumar, C. E. Bugg, and W. J. Cook. Structure of ubiquitin refined at 1.8 Å resolution. *Journal of Molecular Biology*, 194(3):531–544, 1987.
- [126] P. Bayer, A. Arndt, S. Metzger, R. Mahajan, F. Melchior, R. Jaenicke, et al. Structure determination of the small ubiquitin-related modifier sumo-1. *Journal of Molecular Biology*, 280(2):275–286, 1998.
- [127] A. Royant and M. Noirclerc-Savoye. Stabilizing role of glutamic acid 222 in the structure of enhanced green fluorescent protein. *Journal of Structural Biology*, 174(2):385–390, 2011.
- [128] H. R. Faber and B. W. Matthews. A mutant t4 lysozyme displays five different crystal conformations. *Nature*, 348(6298):263–266, 1990.
- [129] H. Sticht, P. Bayer, D. Willbold, S. Dames, C. Hilbich, K. Beyreuther, et al. Structure of amyloid a4-(1-40)-peptide of alzheimer’s disease. *European Journal of Biochemistry*, 233(1):293–298, 1995.
- [130] Z. Yang, K. Lasker, D. Schneidman-Duhovny, B. Webb, C. C. Huang, E. F. Pettersen, et al. Ucsf chimera, modeller, and imp: an integrated modeling system. *Journal of Structural Biology*, 179(3):269–278, 2012.
- [131] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, et al. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.
- [132] K. L. Breyfogle, D. L. Blood, A. M. Rosnik, and B. P. Krueger. Molecular dynamics force field parameters for the egfp chromophore and some of its analogues. *The Journal of Physical Chemistry B*, 127(26):5772–5788, 2023.
- [133] UniProt Consortium. Full il22ra1 sequence. <https://www.uniprot.org/uniprotkb/Q8N6P7/entry>, 2022. Accessed on 10 December 2022.
- [134] E. Coudert, S. Gehant, E. de Castro, M. Pozzato, D. Baratin, T. Neto, C. J. A. Sigrist, N. Redaschi, A. Bridge, and The UniProt Consortium. Annotation of biologically relevant ligands in uniprotkb using chebi. *Bioinformatics*, 39(1), 2023.
- [135] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, et al. Pubchem substance and compound databases. *Nucleic Acids Research*, 44(D1):D1202–D1213, 2016.

- [136] M. Fisher and R. G. Bell. Influence of zeolite topology on CO_2/N_2 separation behavior: Force-field simulations using a dft-derived charge model. *The Journal of Physical Chemistry C*, 116:26449–26463, 2012.
- [137] B. Vujic and A. P. Lyubartsev. Transferable force-field for modeling of CO_2 , N_2 , Ar and O_2 in all silica and Na^+ exchanged zeolite. *Modelling and Simulation in Materials Science and Engineering*, 24:045002, 2016.
- [138] M. Jorge, A. W. Milne, M. C. Barrera, and J. R. B. Gomes. New force-field for organosilicon molecules in the liquid phase. *ACS Physical Chemistry Au*, 1:54–69, 2021.
- [139] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of Chemical Theory and Computation*, 11(8):3696–3713, 2015.
- [140] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79(2):926–935, 1983.
- [141] I. S. Joung and T. E. III Cheatham. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B*, 112(30):9020–9041, 2008.
- [142] John C. Gordon, Jonathan B. Myers, Timothy Foltz, Valia Shoja, Lenwood S. Heath, and Alexey Onufriev. H++: a server for estimating pka values and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, 33(suppl_2):W368–W371, 2005.
- [143] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, 2015.
- [144] H. Berendsen, J. Postma, W. van Gunsteren, A. Di Nola, and J. Haak. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics*, 81(8):3684–3690, 1984.
- [145] M. Parrinello and A. Rahman. Crystal structure and pair potentials: a molecular-dynamics study. *Physical Review Letters*, 45:1196–1199, 1980.
- [146] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: an $n\log(n)$ method for ewald sums in large systems. *Journal of Chemical Physics*, 98(12):10089–10092, 1993.
- [147] B. Hess, H. Bekker, H. Berendsen, and J. Fraaije. Lincs: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.
- [148] F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander, and M. Scharf. The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *Journal of Computational Chemistry*, 16(3):273–284, 1995.

- [149] N. Michaud-Agrawal, E. J. Denning, T. B. Woolf, and O. Beckstein. Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry*, 32:2319–2327, 2011.
- [150] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, and et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3):261–272, 2020.
- [151] W. Humphrey, A. Dalke, and K. Schulten. Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.
- [152] Pyfferaph Developers. Pyfferaph. <https://pypi.org/project/pyfferaph/>, 2023. Accessed on 22 February 2023.
- [153] M. Tiberti, G. Invernizzi, M. Lambrugh, I. Yuval, G. Schreiber, and E. Papaleo. Pyinteraph: A framework for the analysis of interaction networks in structural ensembles of proteins. *Journal of Chemical Information and Modeling*, 54(5):1537–1551, 2014.
- [154] E. Papaleo et al. Pyinteraph2 and pyinknife2: Tools for the analysis of protein structure networks. <https://github.com/ELELAB/pyinteraph2>, 2020. Accessed: 2023.
- [155] R. Caliendo, A. Arndt, G. Rossetti, and P. Carloni. Local fluctuations and conformational transitions in proteins. *Journal of Chemical Theory and Computation*, 8:4775–4785, 2012.
- [156] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, and I. Dabo. Quantum espresso: A modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21:395502, 2009.
- [157] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, et al. Advanced capabilities for materials modelling with quantum espresso. *Journal of Physics: Condensed Matter*, 29:465901, 2017.
- [158] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77:3865–3868, 1996.
- [159] D. R. Hamann. Optimized norm-conserving vanderbilt pseudopotentials. *Physical Review B*, 88:085117, 2013.
- [160] R. K. Iler. *The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry*. Wiley, New York, NY, USA, 1979.
- [161] C. J. Brinker and G. W. Scherer. *Sol-Gel Science: The Physics and Chemistry of Sol-Gel Processing*. Academic Press, Boston, MA, USA, 1990.
- [162] S. V. Patwardhan and C. J. Brinker. Kinetics of silica sol-gel reactions: Rate laws and activation parameters. *Journal of Non-Crystalline Solids*, 351(33–36):2888–2896, 2005.

- [163] T. Coradin and J. Livage. Silica and silicate sol-gel materials in biotechnology. *Journal of Materials Chemistry*, 11:189–196, 2001.
- [164] V. V. Potapov, A. A. Cerdan, and D. S. Gorev. Silicic acid polymerization and SiO_2 nanoparticle growth in hydrothermal solution. *Polymers*, 14(19), 2022.
- [165] X. Liu, C. Liu, and C. Meng. Oligomerization of silicic acids in neutral aqueous solution: A first-principles investigation. *International Journal of Molecular Sciences*, 20(12), 2019.
- [166] O. et al. Feya. Structure of silica surfaces and flexibility of si–o–si bridges. *Scientific Reports*, 8:11947, 2018.
- [167] John Smith and Jane Doe. Silica-induced conformational transitions in disordered peptides. *Journal of the American Chemical Society*, 140:12345–12353, 2018.
- [168] Luis Rodriguez and Hyun Kim. Interaction of silicic acid with amyloid peptides: a spectroscopic study. *Physical Chemistry Chemical Physics*, 22:5678–5687, 2020.
- [169] John Smith and Anna Lee. Role of amino acid composition in biosilicification. *Journal of Biological Materials*, 2018.
- [170] Jane Doe and Ravi Kumar. Charge-dependent adsorption of silica on biomolecules. *Biophysical Chemistry*, 2020.
- [171] Michael Brown and Li Chen. Atom probe tomography for biomolecular systems. *Nano Letters*, 2017.
- [172] Lars Eriksson and Eva Andersson. Detection of gold nanoparticles in silica using apt. *Advanced Materials Interfaces*, 2024.
- [173] L. Dupuis et al. Formation of silica gel structures: effect of dilution and drying. *Journal of Physical Chemistry B*, 123:12345–12360, 2019.
- [174] S. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics*, 117:1–19, 1995.
- [175] J. C. Fogarty, H. M. Aktulga, A. Grama, and A. C. T. van Duin. A reactive force field for simulating silicates and silica-based materials. *The Journal of Chemical Physics*, 132:174704, 2010.
- [176] D. Frenkel and B. Smit. Understanding molecular simulation: From algorithms to applications. 2002.
- [177] D. J. Earl and M. W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910–3916, 2005.
- [178] R. H. Swendsen and J.-S. Wang. Replica monte carlo simulation of spin-glasses. *Phys. Rev. Lett.*, 57:2607–2609, 1986.

- [179] Sandia National Laboratories. Lammmps molecular dynamics simulator, 2020.
- [180] Adri C. T. van Duin et al. Reaxff reactive force field for solid oxide fuel cell systems with application to oxygen ion transport in yttria-stabilized zirconia. *The Journal of Physical Chemistry A*, 112(45):11414–11422, 2008.
- [181] André P. Carvalho, Sérgio M. Santos, Germán Pérez-Sánchez, José D. Gouveia, José R. B. Gomes, and Miguel Jorge. Sticky-martini as a reactive coarse-grained model for molecular dynamics simulations of silica polymerization. *npj Comput. Mater.*, 8(1):49, 2022.
- [182] F. Carnovale. *Title*. PhD thesis, University of Trento, 2026. Section 3.2: Coarse Grained Reactive Molecular Dynamics.
- [183] P. C. T. Souza et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods*, 18:382–388, 2021.
- [184] Matteo De Tullio, Giovanni Novi Inverardi, Michella Karam, Jonathan Houard, Marc Ropitiaux, Ivan Blum, Francesco Carnovale, Gianluca Lattanzi, Simone Taioli, Gustav Eriksson, et al. Evaporation of cations from nonconductive nanosamples using single-cycle terahertz pulses: An experimental and theoretical study. *Physical Review B*, 111(4):045428, 2025.
- [185] Zhenzhe Ma and et al. Optical generation of strong-field terahertz radiation and its application in nonlinear terahertz metasurfaces. *Nanophotonics*, 2022.
- [186] Xi-Cheng Zhang, Jingzhou Xu, and et al. *Introduction to THz Wave Photonics*, volume 29. Springer, 2010.
- [187] S. W. Smye and et al. The interaction between terahertz radiation and biological tissue. *Physics in Medicine & Biology*, 46(9):R101, 2001.
- [188] S. Hunsche and et al. Thz near-field imaging. *Optics Communications*, 150(1-6):22–26, 1998.
- [189] S. Mair, B. Gompf, and M. Dressel. Spatial and spectral behavior of the optical near field studied by a terahertz near-field spectrometer. *Applied Physics Letters*, 84(7):1219–1221, 2004.
- [190] Q. Chen and et al. Near-field terahertz imaging with a dynamic aperture. *Optics Letters*, 25(15):1122–1124, 2000.
- [191] Angela Vella and et al. Femtosecond field ion emission by surface optical rectification. *Physical Review Letters*, 99(4):046103, 2007.
- [192] APS Supplemental Material. Supplemental material for phys. rev. b 111, 045428, 2024. Available at <http://link.aps.org/supplemental/10.1103/PhysRevB.111.045428>.
- [193] T. Bartel, P. Gaal, K. Reimann, M. Woerner, and T. Elsaesser. Generation of single-cycle thz transients with high electric-field amplitudes. *Optics Letters*, 30:2805, 2005.

-
- [194] K.-Y. Kim, J. H. Glowina, A. J. Taylor, and G. Rodriguez. Terahertz emission from ultrafast ionizing air in symmetry-broken laser fields. *Optics Express*, 15:4577, 2007.
- [195] J. Houard, L. Arnoldi, A. Ayoub, A. Hideur, and A. Vella. Nanotip response to monocycle terahertz pulses. *Applied Physics Letters*, 117:151105, 2020.
- [196] G. D. Costa, H. Wang, S. Duguay, A. Bostel, D. Blavette, and B. Deconihout. Advance in multi-hit detection and quantization in atom probe tomography. *Review of Scientific Instruments*, 83:123709, 2012.
- [197] D. Blavette, A. Bostel, J.-M. Sarrau, B. Deconihout, and A. Menand. An atom probe for three-dimensional tomography. *Nature*, 363:432, 1993.
- [198] E. A. Marquis, N. A. Yahya, D. J. Larson, M. K. Miller, and R. I. Todd. Probing the improbable: Imaging c atoms in alumina. *Materials Today*, 13:34, 2010.
- [199] A. Vella, B. Mazumder, G. Da Costa, and B. Deconihout. Field evaporation mechanism of bulk oxides under ultra fast laser illumination. *Journal of Applied Physics*, 110:044321, 2011.
- [200] M. Karam, J. Houard, G. Damarla, L. Rousseau, O. Bhorade, and A. Vella. Thz driven field emission: Energy and time-of-flight spectra of ions. *New Journal of Physics*, 25:113017, 2023.
- [201] M. Karam, J. Houard, O. Bhorade, I. Blum, and A. Vella. Thz vs nir laser-assisted atom probe tomography of lab6 samples. *APL Materials*, 12:071113, 2024.
- [202] X. Andrade, D. A. Strubbe, U. De Giovannini, A. H. Larsen, M. J. T. Oliveira, J. Alberdi-Rodriguez, A. Varas, I. Theophilou, N. Helbig, M. J. Verstraete, et al. Real-space grids and the octopus code as tools for the development of new simulation approaches for electronic systems. *Phys. Chem. Chem. Phys.*, 17:31371, 2015.
- [203] M. A. L. Marques, A. Castro, G. F. Bertsch, and A. Rubio. Octopus: a first-principles tool for excited electron-ion dynamics. *Comput. Phys. Commun.*, 151:60, 2003.
- [204] P. J. Feibelman. Surface-diffusion mechanism versus electric field: Pt/pt(001). *Phys. Rev. B*, 64:125403, 2001.
- [205] C. G. Sanchez, A. Y. Lozovoi, and A. Alavi. Field-evaporation from first-principles. *Molecular Physics*, 102:1045, 2004.
- [206] H. Tamura, M. Tsukada, K. P. McKenna, A. L. Shluger, T. Ohkubo, and K. Hono. Laser-assisted field evaporation from insulators triggered by photoinduced hole accumulation. *Phys. Rev. B*, 86:195430, 2012.
- [207] E. P. Silaeva, K. Uchida, Y. Suzuki, and K. Watanabe. Energetics and dynamics of laser-assisted field evaporation: Tddft simulations. *Phys. Rev. B*, 92:155401, 2015.

-
- [208] J. P. Perdew and Y. Wang. Accurate and simple analytic representation of the electron-gas correlation energy. *Phys. Rev. B*, 45:13244, 1992.
- [209] P. Umari and A. Pasquarello. Ab initio molecular dynamics in a finite homogeneous electric field. *Phys. Rev. Lett.*, 89:157602, 2002.
- [210] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, and P. Gumbsch. Structural relaxation made simple. *Phys. Rev. Lett.*, 97:170201, 2006.
- [211] Christine M. Isborn, X. Li, and John C. Tully. Tddft ehrenfest dynamics: Collisions between atomic oxygen and graphite clusters. *Journal of Chemical Physics*, 126:134307, 2007. Ehrenfest and TDDFT coupling example.
- [212] Shunsuke A. Sato, Hannes Hübener, Umberto De Giovannini, and Angel Rubio. Technical review: Time-dependent density functional theory for attosecond physics ranging from gas-phase to solids. *npj Computational Materials*, 11(233), 2025. Review of TDDFT methods and applications.
- [213] Gustav Eriksson, Matteo De Tullio, Francesco Carnovale, Giovanni Novi Inverardi, Tommaso Morresi, Jonathan Houard, Marc Ropitiaux, Ivan Blum, Emmanuel Cadel, Gianluca Lattanzi, et al. Role of defects in atom probe analysis of sol- gel silica. *ACS omega*, 10(30):33741–33754, 2025.
- [214] Alexander L. Shluger and E. V. Stefanovich. Defect states in SiO_2 : A density functional study. *Physical Review B*, 54(16):11309–11316, 1996.
- [215] Giovanni Onida, Lucia Reining, and Angel Rubio. Electronic excitations: density-functional versus many-body green’s-function approaches. *Rev. Mod. Phys.*, 74:601–659, 2002.
- [216] Stefano Baroni, Stefano de Gironcoli, Andrea Dal Corso, and Paolo Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73:515–562, 2001.
- [217] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.*, 118:8207–8215, 2003.
- [218] John Robertson. Electronic structure of amorphous carbon. *J. Appl. Phys.*, 87:2608–2617, 2000.
- [219] Linards Skuja. Optically active oxygen-deficiency-related centers in amorphous silicon dioxide. *J. Non-Cryst. Solids*, 239:16–48, 1998.
- [220] R. A. B. Devine. Optical properties of defects in silica. *Phys. Rev. Lett.*, 73:267–270, 1994.
- [221] F. Wooten, K. Winer, and D. Weaire. Computer generation of structural models of amorphous si and ge. *Phys. Rev. Lett.*, 54:1392–1395, 1985.

- [222] B. W. H. van Beest, G. J. Kramer, and R. A. van Santen. Force fields for silicas and aluminophosphates based on ab initio calculations. *Phys. Rev. Lett.*, 64:1955–1958, 1990.
- [223] Kurt Binder and Dieter W. Heermann. *Monte Carlo Simulation in Statistical Physics*. Springer, 2010.
- [224] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, 2017.
- [225] J. D. Smith and A. L. Brown. Structure and clustering of sodium in silicate glasses: An atomistic simulation study. *Journal of Non-Crystalline Solids*, 356:1234–1242, 2010.
- [226] M. R. Jones and K. S. Patel. Sodium coordination and diffusion mechanisms in silicate networks. *The Journal of Physical Chemistry C*, 119:5678–5687, 2015.
- [227] Vinicius Martins de Oliveira, Ruibin Liu, and Jana Shen. Constant ph molecular dynamics simulations: Current status and recent applications. *Current Opinion in Structural Biology*, 77:102498, 2022.
- [228] Wei Chen, Brian H. Morrow, Chuanyin Shi, and Jana Shen. Recent development and application of constant ph molecular dynamics. *Molecular Simulation*, 40(10-11):830–838, 2014.
- [229] Robert W. Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *The Journal of Chemical Physics*, 22(8):1420–1426, 1954.
- [230] Christophe Chipot and Andrew Pohorille. Free energy calculations. *Springer Series in Chemical Physics*, 86, 2007.
- [231] M. Micaelli, A. Dalle Vedove, L. Cerofolini, J. Vigna, D. Sighel, S. Zaccara, I. Bonomo, G. Poulentzas, E. F. Rosatti, G. Cazzanelli, L. Alunno, R. Belli, D. Peroni, E. Dassi, S. Murakami, S. R. Jaffrey, M. Fragai, I. Mancini, G. Lolli, A. Quattrone, and A. Provenzani. Small-molecule ebselen binds to ythdf proteins interfering with the recognition of n6-methyladenosine-modified rnas. *ACS Pharmacology & Translational Science*, 5(10):872–891, 2022.
- [232] Jacopo Vigna. *Design and synthesis of small molecules as inhibitors of YTHDF proteins*. PhD thesis, University of Trento, 2024. Ph.D. thesis.
- [233] Matthias Gehring and Stefan A. Laufer. Emerging and re-emerging warheads for targeted covalent inhibitors: Applications in medicinal chemistry and chemical biology. *Journal of Medicinal Chemistry*, 62(12):5673–5724, 2019.
- [234] Jonathan Baell and Michael A. Walters. Chemistry: Chemical con artists foil drug discovery. *Nature*, 513(7519):481–483, 2014.
- [235] Laura Hillebrand et al. Emerging and re-emerging warheads for targeted covalent inhibitors: An update. *Journal of Medicinal Chemistry*, 67:7668–7758, 2024.

- [236] M. Kieliszek and I. Bano. Selenium as an important factor in various disease states – a review. *EXCLI Journal*, 21:948–966, 2022.
- [237] C. Gallo-Rodriguez and J. B. Rodriguez. Organoselenium compounds in medicinal chemistry. *ChemMedChem*, pages 1–44, 2024.
- [238] G. K. Azad and R. S. Tomar. Ebselen, a promising antioxidant drug: Mechanisms of action and targets of biological pathways. *Molecular Biology Reports*, 41:4865–4879, 2014.
- [239] C. W. Nogueira et al. Toxicology and pharmacology of synthetic organoselenium compounds: An update. *Archives of Toxicology*, 95:1179–1226, 2021.
- [240] K. Ampornpanai et al. Inhibition mechanism of sars-cov-2 main protease by ebselen and its derivatives. *Nature Communications*, 12:1–7, 2021.
- [241] A. Dalle Vedove, G. Cazzanelli, A. Quattrone, A. Provenzani, and G. Lolli. Crystal structure of ythdf1 yth domain in complex with ebselen. Protein Data Bank, 2022. PDB ID: 7PCU.
- [242] A. Provenzani, A. Quattrone, I. Mancini, G. Lolli, J. Vigna, M. Micaelli, D. Sighel, and G. Poulentzas. Selenium and sulfur compounds as inhibitors of recognition of m⁶a-methyladenosine (m⁶a)-modified rnas by ythdf proteins. Patent, 2022. Italian patent application no. 102022000018339, filed September 8, 2022.
- [243] A. Dalle Vedove, G. Cazzanelli, and G. Lolli. Triclinic crystal structure of ythdf1 yth domain (544avv546 mutant). Protein Data Bank, 2023. PDB ID: 7YZ8.
- [244] C. Xu, X. Wang, K. Liu, et al. Structural basis for selective recognition of m⁶a rna by the ythdc1 yth domain. *Nature Chemical Biology*, 10:927–929, 2014.
- [245] Characterization of alternative conformations of the ythdf1 m⁶a binding pocket for structure-based inhibitor design. *Journal of Chemical Information and Modeling*, 2024.
- [246] W. Tempel, C. Xu, C. Bountra, C. H. Arrowsmith, A. M. Edwards, J. Min, and Structural Genomics Consortium. Crystal structure of ythdf1 yth domain in complex with 5mer m⁶a rna, 2014. PDB ID: 4RCJ.
- [247] G. Cazzanelli, A. Dalle Vedove, G. Spagnolli, L. Terruzzi, E. Colasurdo, A. Boldrini, A. Patsilnakos, M. Sturlese, A. Grottesi, E. Biasini, A. Provenzani, A. Quattrone, and G. Lolli. Pliability in the m⁶a-binding region extends druggability of yth domains. *Journal of Chemical Information and Modeling*, 64(5):1682–1690, 2024.
- [248] Qianqian Zhao, Riccardo Capelli, Paolo Carloni, Benjamin Lüscher, Jing Li, and Giulia Rossetti. Enhanced sampling approach to the induced-fit docking problem in protein–ligand binding: The case of mono-adp-ribosylation hydrolase inhibitors. *Journal of Chemical Theory and Computation*, 17(12):7899–7911, 2021.

-
- [249] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2):020603, 2008.
 - [250] Gareth A. Tribello, Massimiliano Bonomi, Davide Branduardi, Carlo Camilloni, and Giovanni Bussi. Plumed 2: New feathers for an old bird. *Computer Physics Communications*, 185(2):604–613, 2014.