



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

NANOSCIENZE PER LA MEDICINA E PER L'AMBIENTE

Ciclo 38

Settore Concorsuale: 03/A2 – MODELLI E METODOLOGIE PER LE SCIENZE CHIMICHE

Settore Scientifico Disciplinare: CHIM/02 – CHIMICA FISICA

A MULTISCALE COMPUTATIONAL FRAMEWORK FOR THE DESIGN AND
SCREENING OF GRAPHENE OXIDE IN WASTEWATER REMEDIATION

Presentata da: *Alessandro Calza*

Coordinatore Dottorato

Prof. Matteo Calvaresi

Supervisore

Prof. Francesco Zerbetto

Cosupervisore

Prof. Matteo Calvaresi

Esame finale anno 2026

ABSTRACT

Water pollution and freshwater scarcity have emerged as critical global environmental issues with consequences for public health, ecosystem integrity, and sustainable development. Accelerated urbanization, demographic expansion, industrial growth, and intensified agricultural practices have increased freshwater demand and exerted unprecedented stress on global water resources. Within this context, a major and growing concern is the proliferation of Emerging Contaminants (ECs), also referred to as Contaminants of Emerging Concern. These substances, predominantly of anthropogenic origin, are often neither comprehensively regulated nor systematically monitored, despite mounting evidence of their ecological and toxicological relevance.

Representative classes of ECs include pharmaceuticals, personal care products, disinfection by-products, industrial additives, surfactants, per- and polyfluoroalkyl substances (PFAS), pesticides, microplastics, heavy metals, and engineered nanomaterials. Conventional water and wastewater treatment technologies generally exhibit limited removal efficiencies for many of these compounds. Consequently, ECs are frequently detected in the effluents of wastewater treatment plants, as well as in sediments and surface waters located downstream of urban and industrial areas, thereby promoting their recirculation within the hydrological cycle.

In recent years, filtration materials based on graphene oxide (GO) and its derivatives have demonstrated adsorption performances comparable to, or surpassing, those of traditional sorbents, and thus represent a promising strategy for the mitigation of EC pollution. Nevertheless, the rapidly expanding number of ECs, coupled with their pronounced structural and chemical heterogeneity, poses a substantial challenge to the experimental design and optimization of GO-based adsorbents with tailored and sufficiently high pollutant-binding capacities.

To address this limitation, computational approaches, such as virtual screening (VS) and molecular dynamics (MD) simulations, offer an efficient means to evaluate the interactions of thousands of candidate molecules with GO-based materials and to identify those exhibiting the most favorable binding characteristics. These methodologies can significantly accelerate the rational development of novel graphene oxide derivatives with enhanced filtration

performance. In this framework, the present doctoral thesis aims to exploit established computational protocols traditionally employed in biophysics and adapt them to material science applications, thereby establishing a robust multiscale computational pipeline for the systematic investigation of EC@GO interactions.

An *in-house* structural database comprising approximately 2000 emerging contaminants was constructed based on recent literature. One-dimensional molecular descriptors were extracted from existing databases and subsequently converted into three-dimensional representations by explicitly accounting for the correct stereochemistry and protonation states at environmentally relevant pH values. Then, a scalable graphene oxide model was developed on the basis of experimental data and commercially available LayerOne products, yielding a library of GO structures with varying degrees of oxidation.

A virtual screening campaign, followed by all-atom molecular dynamics (AA-MD) simulations, was performed to comprehensively sample the material surface and to quantify and rank the interactions within each EC@GO pair. Specifically, the protocol employed a grid-based docking strategy to ensure exhaustive exploration of the graphene oxide surface topology, generating more than 86000 initial poses, which were subsequently refined through cumulative MD simulations totaling approximately 29 microseconds.

Binding free energies were computed using the MM-GBSA method combined with the GB-OBC2 implicit solvation model and further coupled with Normal Mode (NMODE) analysis to explicitly account for entropic contributions. The results indicate that adsorption is predominantly controlled by the relative orientation of the adsorbate on the heterogeneous GO surface rather than by its stereochemical configuration, as demonstrated by the largely conserved binding profiles across distinct stereoisomers. Inclusion of entropic terms proved essential to correct the size bias intrinsic to enthalpy-only scoring, revealing that high affinity ligands such as atorvastatin and bis(2-ethylhexyl) phthalate incur substantial thermodynamic penalties due to conformational rigidification upon adsorption.

The final thermodynamic ranking on highly oxidized graphene oxide (GOX8-P935) indicates a predominance of hydrophobic and stacking interactions, with polycyclic aromatic hydrocarbons and plasticizers displaying the strongest affinities, while polar and anionic species are disfavored due to pronounced de-

solvation penalties. Expanding this protocol to reduced graphene oxide (REGOX8-P980) reveals an enhanced binding affinity that is almost exclusively enthalpy-driven. The restoration of planar basal carbon domains not only maximizes surface contact for neutral molecules but also stabilizes specific positively charged drugs, such as azithromycin and propranolol, by providing geometrically unhindered binding pockets. However, the counterintuitive preference of positive compounds for the less electronegative REGOX8-P980 surface exposes the inherent limitations of representing the nanomaterial through a single random Lerf-Klinowski microstate. Consequently, a methodological evaluation demonstrates that reliable *in silico* affinity predictions require a more robust approach: energy minimization of static complexes fails to capture dynamic surface nesting; Quasi-Harmonic entropy calculations demand extensive sampling to match NMODE accuracy; and future computational models must average thermodynamic results across an ensemble of heterogeneous structural replicas to accurately reflect the macroscopic experimental reality.

In a complementary effort, a library of linear PFAS molecules was parameterized using the Martini 3 coarse grain force field, by reproducing both their predicted and experimental physicochemical behavior. This coarse-grained database was subsequently employed to simulate: (i) homogeneous solutions of PFOA and PFDoDA; (ii) a mixture of PFAS; and (iii) a mixture of emerging contaminants. For systems (ii) and (iii), simulations were carried out both in the presence and absence of graphene to investigate differential adsorption phenomena under conditions that approximate realistic environmental scenarios, providing mechanistic insights into competitive adsorption processes that would be prohibitively expensive to obtain using conventional all-atom molecular dynamics alone.

We acknowledge the ISCRA initiative for awarding this project access to the LEONARDO supercomputer owned by the EuroHPC Joint Undertaking, hosted by CINECA (Italy). Alessandro Calza is funded by the European Union – Next Generation EU under the Italian National Recovery and Resilience Plan (PNRR M4C2, Investment 1.4, Notice n. 3138 of 16/12/2021, CN00000013, CUP J33C22001170001), as part of the ICSC SPOKE 6 “Multiscale Modelling and Engineering Applications”, Flagship Project FP1 “Advanced Materials and New Devices” – WP1: Simulations of Complex Multiscale and Multiphysics Models.

CONTENTS

<u>ABSTRACT</u>	<u>I</u>
<u>FIGURE INDEX</u>	<u>VI</u>
<u>TABLE INDEX</u>	<u>VII</u>
<u>LIST OF ACRONYMS</u>	<u>VIII</u>
<u>CHAPTER 1: INTRODUCTION</u>	<u>1</u>
1.1 EMERGING CONTAMINANTS	1
1.2 CARBON NANOMATERIALS	6
1.3 GRAPHENE OXIDE IN WASTEWATER REMEDIATION	7
1.4 AIM OF THE THESIS	10
<u>CHAPTER 2: DATABASE DESIGN</u>	<u>12</u>
2.1 METHODS	12
2.1.1 THE PUBCHEM DATABASE	12
2.1.2 MOLECULE STRING NOTATIONS	13
2.1.3 QUANTUM MECHANICS (QM) AND AMBER PARAMETRIZATION	17
2.2 EMERGING CONTAMINANTS DATABASE (EC-DB)	20
2.3 GRAPHENE OXIDE DATABASE (GO-DB)	23
<u>CHAPTER 3: VIRTUAL SCREENING</u>	<u>27</u>
3.1 METHODS	27
3.1.1 VIRTUAL SCREENING AND DOCKING	27
3.1.2 MOLECULAR DYNAMICS	34
3.1.3 MM-GBSA/MM-PBSA METHODS	39
3.2 GRID-BASED DOCKING AND ALL-ATOM MD	46
3.3 RESULTS	49
3.3.1 RATIONALE FOR SELECTION OF BINDING AFFINITY PROTOCOL	49
3.3.2 FINAL RANKING AND FOCUS ON SELECTED CONTAMINANTS	56
3.3.3 COMPARISON WITH REGOX8-P980	70
3.3.4 SAR ANALYSIS AND TOPOLOGICAL PROFILING	73

3.3.5	PROTOCOL OPTIMIZATION	78
CHAPTER 4: COARSE-GRAINING		84
4.1	METHODS	85
4.2	MARTINI 3 PFAS MAPPING AND PARAMETRIZATION	87
4.3	MODEL VALIDATION	91
4.3.1	SASA AND FREE ENERGIES OF TRANSFER	91
4.3.2	ADSORPTION ON GRAPHENE	93
4.3.3	PHASE BEHAVIOR	95
4.4	COMPLEX POLLUTANTS MIXTURE SIMULATIONS	102
4.5	APPENDIX	106
CHAPTER 5: CONCLUSIONS		119
5.1	LIMITATIONS AND FUTURE PERSPECTIVES	121
REFERENCES		125

FIGURE INDEX

FIGURE 1: origin and transport of Emerging Contaminants (ECs).	5
FIGURE 2: graphene oxide structures and PSU-GO fibers.	10
FIGURE 3: aim of the thesis.	11
FIGURE 4: structural identifiers of ibuprofen.	16
FIGURE 5: workflow for the development of the emerging contaminants database (EC-DB).	23
FIGURE 6: parametrization of graphene oxide functional groups.	24
FIGURE 7: workflow for the development of the layerone graphene oxide database (GO-DB).	26
FIGURE 8: molecular docking and virtual screening process.	29
FIGURE 9: docking process with Autodock Vina.	33
FIGURE 10: MM-GBSA thermodynamic cycle.	42
FIGURE 11: grid-based docking strategy for graphene oxide surface exploration.	47
FIGURE 12: comparison of binding affinities among stereoisomers.	50
FIGURE 13: impact of conformational flexibility and steric bulk on binding affinity.	51
FIGURE 14: impact of molecular weight and lipophilicity on ligand efficiency.	52
FIGURE 15: impact of GB solvation models on binding affinity rankings.	55
FIGURE 16: binding free energy ranking of the screened priority contaminants.	57
FIGURE 17: top 15 binders with individual stereoisomers.	60
FIGURE 18: binding analysis of bis(2-ethylhexyl) phthalate (DEHP).	61
FIGURE 19: interaction fingerprint of the DEHP@GO complex.	62
FIGURE 20: binding analysis of atorvastatin.	64
FIGURE 21: interaction fingerprint of the atorvastatin@GO complex.	66
FIGURE 22: binding analysis of chlorpyrifos.	68
FIGURE 23: interaction fingerprint of the chlorpyrifos@GO complex.	69
FIGURE 24: influence of graphene oxide reduction on binding thermodynamics.	70
FIGURE 25: structural impact of surface functionalization on graphene oxide topology.	71
FIGURE 26: impact of surface topology on the binding geometry and energetics of azithromycin and propranolol.	72
FIGURE 27: 2D visualization of the structural diversity within the pollutant dataset.	74
FIGURE 28: average binding free energies across the identified chemical families on the GOX8-P935 surface.	75
FIGURE 29: comparative SAR analysis of the adsorption hierarchy on oxidized versus reduced graphene surfaces.	77
FIGURE 30: comparison between trajectory-averaged and post-docking minimization binding affinities.	79
FIGURE 31: pose agreement percentage across different GB solvation models.	80
FIGURE 32: impact of surface flexibility on ligand binding conformations.	81
FIGURE 33: impact of trajectory sampling density on quasi-harmonic entropy estimates.	82
FIGURE 34: coarse-grained parametrization protocol for Martini 3 models.	86
FIGURE 35: 2D representation of the PFAS considered in the present chapter.	87
FIGURE 36: breakdown of PFOA and PFOS into the three portions used for the CG mapping and their Martini 3 representation.	88
FIGURE 37: comparison of the bonded distributions between the mapped AA reference and the final CG trajectory in case of PFOA.	90
FIGURE 38: comparison of PFOA conolly surface and PFCAs logD values.	92
FIGURE 39: PMF of PFOA desorption from graphene.	94
FIGURE 40: the number of clusters at increasing PFOA concentrations.	95
FIGURE 41: the number of clusters at increasing PFOA concentrations with labeled e and r beads.	98

FIGURE 42: the number of clusters at increasing PFOA concentrations with the addition of NaCl 0.25 M.	99
FIGURE 43: temporal evolution of PFOA and PFDoDA aggregates.	102
FIGURE 44: temporal evolution of the PFAS mixture in the absence and presence of graphene.	104
FIGURE 45: temporal evolution of the emerging contaminants mixture in the absence and presence of graphene.	105
FIGURE 46: the bonded distribution of PFBA (perfluorobutanoic acid).	107
FIGURE 47: the bonded distribution of PFPeA (perfluoropentanoic acid).	108
FIGURE 48: the bonded distribution of PFHxA (perfluorohexanoic acid).	108
FIGURE 49: the bonded distribution of PFHpA (perfluoroheptanoic acid).	109
FIGURE 50: the bonded distribution of PFNA (perfluorononanoic acid).	109
FIGURE 51: the bonded distribution of PFDA (perfluorodecanoic acid).	109
FIGURE 52: the bonded distribution of PFUnDA (perfluoroundecanoic acid).	110
FIGURE 53: the bonded distribution of PFDoDA (perfluorododecanoic acid).	110
FIGURE 54: the bonded distribution of PFBS (perfluorobutanesulfonic acid).	110
FIGURE 55: the bonded distribution of PFPeS (perfluoropentanesulfonic acid).	111
FIGURE 56: the bonded distribution of PFHxS (perfluorohexanesulfonic acid).	111
FIGURE 57: the bonded distribution of PFHpS (perfluoroheptanesulfonic acid).	112
FIGURE 58: the bonded distribution of PFOS (perfluorooctanesulfonic acid).	112
FIGURE 59: the bonded distribution of PFNS (perfluorononanesulfonic acid).	112
FIGURE 60: the bonded distribution of PFDS (perfluorodecanesulfonic acid).	113
FIGURE 61: the bonded distribution of PFUnDS (perfluoroundecanesulfonic acid).	113
FIGURE 62: the bonded distribution of PFDoDS (perfluorododecanesulfonic acid).	113
FIGURE 63: the average SASA distribution of PFCAs.	113
FIGURE 64: the average SASA distribution of PFSAs.	114
FIGURE 65: PFM CURVES OF PFCAS AND PFSAS.	114
FIGURE 66: comparison of PFSAs logD values.	114
FIGURE 67: molecular electrostatic potential maps (MEP) OF PFCAs compounds.	115
FIGURE 68: molecular electrostatic potential maps (MEP) OF PFSAs compounds.	115
FIGURE 69: comparison of PMF curves following the addition of the e electron-rich label to PFOA beads.	116
FIGURE 70: cluster size distribution for each studied concentration (X beads not labeled) over the last 1 μ s.	116
FIGURE 71: cluster size distribution for each studied concentration (X beads + r/e label) over the last 1 μ s.	117
FIGURE 72: cluster size distribution for each studied concentration (+ NaCl 0.25 m) over the last 1 μ s.	117
FIGURE 73: comparison of the RDF, between the center of mass of each PFOA molecule, for every studied system over the last 1 μ s.	118

TABLE INDEX

TABLE 1: parameters for isomer generation.	21
TABLE 2: intra-class thermodynamic variance within the organophosphate insecticides cluster.	76
TABLE 3: the average aggregation number (N_{AGG}), the mean residence time of a surfactant monomer within the micellar aggregate (τ), and the resulting monomer exchange rate (μs^{-1} micelle $^{-1}$) for PFOA systems at different concentrations.	98
TABLE 4: the parameters used in estimating the ζ -potential.	101
TABLE 5: composition of the pollutants mixtures system.	103
TABLE 6: all the considered system in modelling the phase behavior.	106

LIST OF ACRONYMS

A

AA	All-Atom
AA-MD	All-Atom Molecular Dynamics
AC	Activated Carbon
AChE	Acetylcholinesterase
ADT	AutoDock Tools
APIs	Active Pharmaceutical Ingredients
ASCII	American Standard Code for Information Interchange

B

bGO	Brodie-type Graphene Oxide
BO	Born-Oppenheimer

C

CCL	Contaminant Candidate List
CECs	Contaminants of Emerging Concern
CG	Coarse-Grained
CID	Compound Identifier
CIP	Cahn-Ingold-Prelog
CMC	Critical Micelle Concentration
CNTs	Carbon Nanotubes
COM	Center of Mass
CPP	Critical Packing Parameter
Cryo-EM	Cryogenic Electron Microscopy
CVD	Chemical Vapor Deposition

D

DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DEHP	Bis(2-ethylhexyl) phthalate

E

ECs	Emerging Contaminants
EC-DB	Emerging Contaminants Database
EDC	Endocrine-Disrupting Chemical
EPA	Environmental Protection Agency
EQS	Environmental Quality Standards
ESP	Electrostatic Potential

F

FDA	Food and Drug Administration
FF	Force Field
FGs	Functional Groups

G

GAC	Granular Activated Carbon
GAFF	General AMBER Force Field
GB	Generalized Born
GOX/GO	Graphene Oxide
GO-DB	Graphene Oxide Database
GRMs	Graphene-Related Materials
GTOs	Gaussian-Type Orbitals

H

HCT	Hawkins-Cramer-Truhlar
HF-SCF	Hartree-Fock Self-Consistent Field
hGO	Hummers-type Graphene Oxide
HMG-CoA	3-hydroxy-3-methylglutaryl coenzyme A

I

IARC	International Agency for Research on Cancer
InChI	International Chemical Identifier

LIST OF ACRONYMS

ISOF-CNR Istituto per la Sintesi Organica e la Fotoreattività -
Consiglio Nazionale delle Ricerche

IUPAC International Union of Pure and Applied Chemistry

L

LBVS Ligand-Based Virtual Screening

LCAO Linear Combination of Atomic Orbitals

LE Ligand Efficiency

LJ Lennard-Jones

M

MAD Martini Database

MD Molecular Dynamics

MDS Multidimensional Scaling

MEHP Mono(2-ethylhexyl) phthalate

MEP Molecular Electrostatic Potential Map

MK Merz-Kollman

MM Molecular Mechanics

MM-GBSA Molecular Mechanics-Generalized Born Surface Area

MM-PBSA Molecular Mechanics-Poisson Boltzmann Surface Area

μ VT Grand Canonical Ensemble

N

NaPFO Sodium Perfluorooctanoate

NIH National Institutes of Health

NMA/NMODE Normal Mode Analysis

NMR Nuclear Magnetic Resonance

NPT Isothermal-Isobaric Ensemble

NVE Microcanonical Ensemble

NVT Canonical Ensemble

O

OBC Onufriev-Bashford-Case

OPLS-AA Optimized Potentials for Liquid Simulations All-Atom

P

PAHs	Polycyclic Aromatic Hydrocarbons
PB	Poisson-Boltzmann
PBC	Periodic Boundary Conditions
PEG	Polyethylene Glycol
PES	Potential Energy Surface
PFAAs	Perfluoroalkyl Acids
PFAS	Per- and Polyfluoroalkyl Substances
PFBA	Perfluorobutanoic Acid
PFBS	Perfluorobutanesulfonic Acid
PFCAs	Perfluorocarboxylic Acids
PFDA	Perfluorodecanoic Acid
PFDS	Perfluorodecanesulfonic Acid
PFDODA	Perfluorododecanoic Acid
PFDODS	Perfluorododecanesulfonic Acid
PFHpA	Perfluoroheptanoic Acid
PFHpS	Perfluoroheptanesulfonic Acid
PFHxA	Perfluorohexanoic Acid
PFHxS	Perfluorohexanesulfonic Acid
PFNA	Perfluorononanoic Acid
PFNS	Perfluorononanesulfonic Acid
PFOA	Perfluorooctanoic Acid
PFOS	Perfluorooctanesulfonic Acid
PFPeA	Perfluoropentanoic Acid
PFPeS	Perfluoropentanesulfonic Acid
PFSAs	Perfluorosulfonic Acids
PFUnDA	Perfluoroundecanoic Acid
PFUnDS	Perfluoroundecanesulfonic Acid
PME	Particle Mesh Ewald
PMF	Potential of Mean Force
PPAR	Peroxisome Proliferator-Activated Receptor
PPCPs	Pharmaceuticals and Personal Care Products
PSU	Polysulfone

LIST OF ACRONYMS

PVC Polyvinyl Chloride

Q

QH Quasi-Harmonic

QM Quantum Mechanics

QSAR Quantitative Structure-Activity Relationship

QSPR Quantitative Structure-Property Relationships

R

RC Reaction Coordinate

RDF Radial Distribution Function

REACH Registration, Evaluation, Authorization and Restriction of Chemicals

REGOX/rGO Reduced Graphene Oxide

REMD Replica Exchange Molecular Dynamics

RESP Restrained Electrostatic Potential

S

SAR Structure-Activity Relationship

SASA Solvent Accessible Surface Area

SBVS Structure-Based Virtual Screening

SHA Secure Hash Algorithm

SI Supporting Information

SMILES Simplified Molecular Input Line Entry System

STOs Slater-Type Orbitals

SVHC Substance of Very High Concern

U

UFF Universal Force Field

US Umbrella Sampling

V

VdW van der Waals

VS Virtual Screening

W

WFD Water Framework Directive

WHAM Weighted Histogram Analysis Method

WWTPs Wastewater Treatment Plants

X

XPS X-ray Photoelectron Spectroscopy

CHAPTER 1: INTRODUCTION

This dissertation is structured into five chapters, organized as follows.

Chapter 1 introduces the current state of research on emerging contaminants (Subsection 1.1). Subsections 1.2 and 1.3 provide an overview of the physicochemical properties and technological applications of carbon nanomaterials and graphene oxide, with particular emphasis on their use as adsorbent materials for wastewater treatment. The specific objectives of the thesis are defined in Subsection 1.4.

Chapter 2 details the methodologies adopted for the development of the pollutant database (Subsection 2.2) and the graphene oxide database (Subsection 2.3), hereafter referred to as EC-DB and GO-DB, respectively. These databases constitute the foundation for the analyses presented in the subsequent chapter.

Chapter 3 describes the virtual screening workflow, which comprises molecular docking, all-atom molecular dynamics simulations, and binding affinity calculations (Subsection 3.3).

Chapter 4 presents all work related to coarse-grained modeling (Subsection 4.2 and 4.3) and to the simulation of complex and realistic environments (Subsection 4.4) using the Martini 3 force field.

Finally, Chapter 5 provides a synthesis of the principal findings of this research. With the exception of the concluding summary chapter, each chapter opens with a concise introductory section that delineates its scope, followed by descriptions of the methodologies and results. A comprehensive list of references is provided at the end of the thesis, following the conclusions.

1.1 Emerging contaminants

Water pollution and freshwater scarcity have emerged as critical global environmental challenges with implications for public health and ecosystem functioning. Rapid urbanization, demographic expansion, industrial growth, and high-intensity agricultural systems have collectively amplified freshwater

demand, imposing unprecedented stress on both surface water and groundwater reservoirs [1,2]. Superimposed on these anthropogenic drivers, climate change further intensifies hydrological stress by modifying precipitation patterns, increasing the frequency, severity, and duration of drought events, and reducing the natural recharge capacity of aquifers [3,4].

Collectively, these trends underscore the need to enhance water quality management and to implement more sustainable methods for wastewater treatment, recovery, and reuse. In response, regulatory bodies, such as the U.S. Environmental Protection Agency (EPA) and the European Union, are progressively reinforcing water governance frameworks by implementing more stringent quality standards and specifying concentration limits, monitoring frequencies, and analytical requirements for an expanding array of contaminants [5–8].

Within this context, increasing scientific attention is being directed toward so-called Emerging Contaminants (ECs), also referred to as Contaminants of Emerging Concern (CECs) [9]. Representative categories encompass pharmaceuticals and personal care products (PPCPs), disinfection by-products, industrial additives, surfactants, per- and polyfluoroalkyl substances (PFAS), pesticides, microplastics, heavy metals, and engineered nanomaterials [8,10–12]. In many instances, emerging contaminants do not represent newly synthesized chemicals; rather, they are known characterized substances for which recent advances in analytical chemistry and toxicological research have revealed environmental and public health implications that were not previously recognized [13]. Although these advances have substantially improved detection limits, facilitated detailed speciation, and enhanced monitoring of micropollutants, considerable knowledge gaps persist with respect to their environmental fate, transformation pathways, mixture effects, and long-term ecological and human health consequences [14].

These substances, predominantly of anthropogenic origin, are often neither comprehensively regulated nor systematically monitored, despite mounting evidence of their ecological and toxicological significance. Typically occurring at trace concentrations in the nanogram to microgram per liter range (ng/L– μ g/L), they are frequently categorized as micropollutants [15]. Nonetheless, many ECs exhibit pronounced environmental persistence, attributable to their chemical stability, bioactivity, and resistance to conventional physicochemical and biological degradation processes. Chronic exposure to low concentrations

of these compounds can elicit subtle yet meaningful responses in plants, animals, and humans, including disruptions of metabolic pathways, endocrine and hormonal dysregulation, and impairments of reproductive performance. Such effects may propagate through aquatic food webs, with the potential to alter community composition and compromise ecosystem structure and function [15–18]. Furthermore, these substances frequently co-occur as complex mixtures and undergo dynamic interactions both among themselves and with dissolved organic matter and resident microbial populations. Consequently, they can induce additive, synergistic, or antagonistic responses that often deviate from those predicted on the basis of single-compound toxicity assays. Robust ecological risk assessment of ECs is therefore methodologically challenging and requires integrative, long-term investigations that explicitly incorporate chronic, complex mixture exposure scenarios [19].

Within the broad spectrum of ECs, pharmaceuticals have attracted particularly intense scientific and regulatory attention. They rank among the most ubiquitous contaminants identified in municipal wastewater effluents and in surface waters. Currently, more than three thousand active pharmaceutical ingredients (APIs) are in use in human and veterinary medicine, including, but not limited to, antibiotics, antivirals, antifungals, antihypertensives, antidepressants, analgesics, and steroidal and non-steroidal hormones. Following administration, many APIs are excreted either partially unmetabolized or as pharmacologically active metabolites. In addition, transformation products formed during biological, chemical, or physical treatment processes (e.g., biotransformation, oxidation, photolysis) may retain biological activity or, in some instances, revert to the parent compound [20]. Persistent pharmaceuticals such as carbamazepine and diclofenac are repeatedly detected at concentrations that approach or exceed proposed or existing guideline values in treated effluents and receiving water bodies [21]. Even at trace concentrations, pharmaceuticals can interfere with endocrine, reproductive, and metabolic pathways in aquatic organisms. Chronic exposure has been shown to modify behavior, impair growth, and alter predator-prey interactions, thereby disrupting community structure, inducing ecosystem imbalances, and promoting the propagation of effects through trophic transfer and bioaccumulation [18].

Antibiotics constitute a particularly critical subcategory of pharmaceutical contaminants because their environmental persistence facilitates the emergence and spread of antimicrobial resistance. Sub-inhibitory concentrations of

antibiotics such as tetracycline and sulfamethoxazole can alter microbial community composition and function, selectively favoring resistant strains and enhancing the horizontal transfer of resistance genes within and among bacterial populations. This process intensifies both environmental and public health impacts associated with antibiotic resistance, reinforcing its recognition as a major global health and ecological challenge [22–24].

Alongside pharmaceuticals, per- and polyfluoroalkyl substances (PFAS) constitute a particularly persistent and hazardous subclass of emerging contaminants. These anthropogenic organofluorine compounds are extensively employed in industrial and commercial sectors owing to their pronounced surface activity, exceptional thermal stability, and high chemical inertness. Such properties underpin their widespread application in surface coatings, food-contact materials, textiles, non-stick formulations, and firefighting foams [25]. Ironically, the same physicochemical characteristics confer extreme environmental persistence, facilitate long-range atmospheric and hydrological transport, and promote bioaccumulation and biomagnification in food chains.

Within the broader PFAS category, perfluoroalkyl acids (PFAAs) are of particular concern because of their pronounced resistance to both abiotic and biotic degradation pathways and their ubiquitous detection across diverse environmental compartments, including surface waters, groundwater, soils, sediments, and biota [26,27]. An expanding body of toxicological and epidemiological evidence has linked PFAS exposure with adverse health outcomes, encompassing developmental toxicity, endocrine-disrupting effects, carcinogenic potential, and immunotoxicity [28]. Emissions from industrial manufacturing processes, inappropriate waste management and disposal practices, as well as diffuse releases from PFAS-containing consumer and industrial products, have collectively resulted in an almost global distribution of these compounds in environmental matrices. This widespread occurrence highlights the need for more stringent regulatory frameworks, alongside the development and implementation of advanced remediation strategies specifically tailored to PFAS removal and destruction [27].

The primary challenge in managing these diverse substances lies within our current sanitation infrastructure. Wastewater treatment plants (WWTPs) represent the core of these facilities and play a pivotal role in the protection of aquatic ecosystems; however, they were not originally designed to achieve efficient removal of ECs. Conventional WWTPs employ sequential physical,

chemical, and biological treatment steps that are primarily optimized for the removal of organic matter, nutrients, suspended solids, and pathogenic microorganisms [29,30]. As a consequence, the inherent design and process configurations of these systems constrain their capacity to effectively eliminate ECs. Emerging pollutants are therefore frequently detected in WWTP effluents, sewage sludge, sediments, and receiving surface waters, which promotes their persistence and recirculation within the hydrological cycle.

In light of these limitations, the retrofitting and upgrading of existing WWTPs with advanced or hybrid treatment technologies has become a central focus of both scientific research and environmental policy. Among the most extensively studied and increasingly implemented approaches are advanced oxidation processes (e.g., ozonation, ultraviolet irradiation, photocatalysis), membrane-based separation technologies, and adsorption processes, particularly those employing activated carbon [31–34]. Despite these technological developments, no single treatment technology currently provides complete, reliable, and broad-spectrum removal of all relevant classes of ECs: the performance of any given process is strongly dependent on the physicochemical characteristics of the target contaminants, the configuration and operational conditions of the treatment system, and local environmental variables, thereby making the conception of a universally applicable and optimally effective treatment solution highly challenging [33,34].

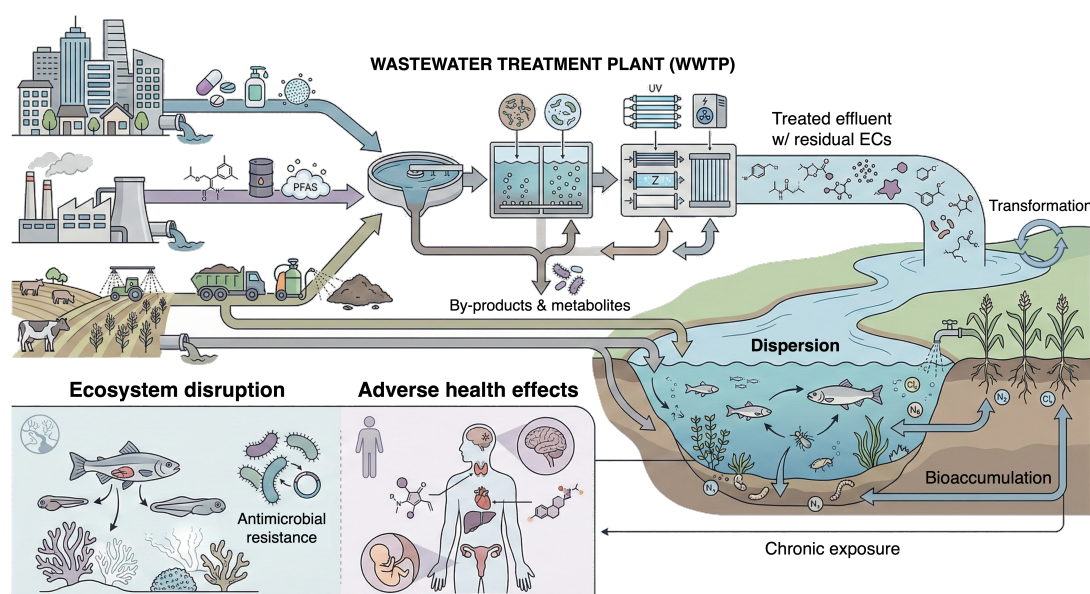


Figure 1: Origin and transport of emerging contaminants (ECs). ECs resist standard wastewater treatments and enter the environment, leading to chronic exposure, bioaccumulation, and cascading effects on ecological and human health.

1.2 Carbon nanomaterials

Over past decades, advances in nanoscience and nanotechnology have fundamentally reshaped strategies for environmental remediation. Significant progress in the design, synthesis, and post-synthetic modification of nanostructured materials has enabled the development of highly engineered platforms to confront escalating global challenges associated with water scarcity and contamination [35]. Nanomaterials have assumed a central role in this technological evolution because their characteristic dimensional regime, typically on the order of 1-100 nm, confers physicochemical properties that differ from those of their bulk analogues [36]. Their exceptionally high surface area to volume ratio provides a dense array of accessible active sites, thereby enhancing performance in adsorption, heterogeneous catalysis, and membrane-based separation processes [37]. In addition, their tunable surface chemistry provides opportunities to engineer materials capable of selectively capturing or degrading trace-level contaminants that are resistant to conventional treatment technologies. The broad range of available synthetic methodologies affords precise control over the morphology, porosity, surface functionality, and physicochemical stability of nanomaterials. Consequently, these systems can be tailored to exhibit optimized pore structures, hydrophilicity/hydrophobicity balance, and thermal or chemical robustness, yielding multifunctional platforms with relevance to catalysis, energy storage, biosensing, and environmental purification [38].

Within the diverse class of nanostructured materials, carbon-based nanomaterials have demonstrated particularly high efficacy owing to their intrinsic chemical stability, versatile surface chemistry, and facile integration into composite architectures with polymers, metals, and metal oxides [39]. Carbon nanomaterials such as activated carbon (AC), carbon nanotubes (CNTs), fullerenes, and graphene are widely employed as adsorbents because of their high specific surface areas, extended π -electron conjugation, and superior mechanical resilience. Although AC has long served as the industrial benchmark for the removal of organic pollutants, novel carbon nanostructures frequently surpass AC in both adsorption capacity and kinetics, largely due to more precisely engineered pore architectures and a higher density and diversity of surface functional groups [40]. These materials can immobilize heavy metals, pharmaceuticals, pesticides, and PFAS through multiple interaction mechanisms, including electrostatic attraction, covalent or coordination bonding, hydrogen bonding, and π - π stacking interactions [41]. In addition, the

intrinsically high electrical conductivity and pronounced chemical inertness of numerous carbon-based materials permit their efficient integration into electrochemical and photocatalytic systems, where they can promote synergistic pollutant degradation while concurrently functioning as active elements in sensing platforms [42].

Within this class of materials, graphene and its derivatives have gained attention owing to their remarkable and frequently unrivaled physicochemical properties. Graphene is composed of a single atomic layer of sp^2 -hybridized carbon atoms organized in a two-dimensional hexagonal honeycomb lattice, constituting the fundamental structural motif of several other carbon allotropes [43]. It is widely recognized as one of the strongest and most versatile materials identified to date, featuring a theoretical specific surface area of approximately $2630 \text{ m}^2 \text{ g}^{-1}$, a Young's modulus on the order of 1 TPa, and a high thermal conductivity exceeding $5000 \text{ W m}^{-1} \text{ K}^{-1}$ [44,45]. Its high optical transmittance ($\sim 98\%$), mechanical flexibility, and exceptional charge carrier mobility further support its application across a broad spectrum of fields, including energy storage and conversion, nanoelectronics, biomedicine, and environmental remediation technologies [46–48].

Synthetic methodologies for graphene-related materials (GRMs) are conventionally classified into top-down and bottom-up routes. In top-down approaches, graphene sheets are generated via the exfoliation of bulk graphite, typically through mechanical, chemical, or electrochemical techniques that disrupt the van der Waals forces responsible for interlayer cohesion in stacked graphene. By contrast, bottom-up strategies assemble graphene architectures from molecular or atomic precursors in the gas or liquid phase, such as methane, ethanol, or aromatic hydrocarbons, most commonly through chemical vapor deposition (CVD), pyrolysis, or related high-temperature processes [49].

1.3 Graphene oxide in wastewater remediation

Among GRMs, graphene oxide (GO) constitutes the most extensively investigated derivative for adsorption-driven purification processes. GO is typically obtained via the oxidative exfoliation of graphite using chemical protocols initially developed by Brodie and subsequently modified by Staudenmaier, Hofmann, and Hummers [50–52]. The resulting material is enriched in oxygenated functional groups, mainly hydroxyl, epoxy, carbonyl, and carboxyl moieties, distributed on both basal planes and sheet edges [53].

These functionalities confer pronounced hydrophilicity and favorable dispersibility in aqueous media, thereby improving compatibility with water-based systems as well as with polymer matrices [54]. Conversely, the introduction of such functional groups substantially diminishes the intrinsic electrical conductivity of the graphene lattice, which can be partially recovered by chemical reduction to yield reduced graphene oxide (rGO) [55–57].

Comparative investigations of GO produced via distinct synthetic routes reveal differences in structure and properties [58–62]. Brodie-type GO (bGO) typically exhibits a lower density of structural defects, a more homogeneous spatial distribution of functional groups, and superior mechanical integrity [63]. In contrast, Hummers-type GO (hGO) generally contains a higher concentration of carboxyl functionalities but also displays enhanced lattice disruption and sheet fragmentation [64]. These structural disparities influence not only the electronic and mechanical behavior but also modulate adsorption capacity and selectivity toward specific contaminants [65].

GO and related materials function as highly efficient sorbents through a combination of concurrent interaction mechanisms. The oxygenated surface groups impart negative surface charge over a broad pH range, which promotes electrostatic attraction toward cationic species, including heavy metal ions and cationic dyes [66]. These functional groups also serve as coordination and binding sites enabling hydrogen bonding and metal ion complexation, whereas the residual sp^2 -hybridized aromatic domains facilitate π - π stacking and van der Waals interactions with aromatic contaminants, such as pharmaceuticals, personal care product residues, and polycyclic aromatic hydrocarbons [67–71]. The coexistence of these multiple binding pathways underlies the high adsorption capacities reported for GO-based sorbents across a wide variety of pollutants, encompassing organic contaminants (e.g., PAHs, halogenated solvents, plasticizers, and endocrine-disrupting hormones), inorganic ions, and engineered or incidental nanoparticles [72–74].

The introduction of additional functionalities via the formation of hybrid nanostructures has further broadened the application space of GO-based adsorbents. Metal and metal oxide nanoparticles, including TiO_2 , ZnO , and Fe_3O_4 , can be immobilized on GO surfaces to impart photoactivity, magnetic responsiveness, or catalytic properties. Such hybrid systems enable combined adsorption and in situ degradation of persistent organic pollutants under light irradiation, thereby offering more sustainable alternatives to conventional

treatment based solely on separation [75–78]. Nonetheless, important questions remain regarding the environmental fate, long term stability, recyclability, and potential nanotoxicological impacts of GO and its hybrids, which are active areas of ongoing research [79–81].

A particularly important recent advancement is the incorporation of graphene oxide into polymeric matrices to produce mechanically robust, reusable, and easily processable composite adsorbents. These polymer-GO systems can be engineered to exhibit tunable porosity, well-defined surface chemistry, and enhanced structural integrity, thereby enabling efficient recovery and regeneration. The resulting combination of properties makes these materials suitable for both large scale wastewater treatment applications and decentralized or point of use water purification systems [61,82–84]. The incorporation of GO into polysulfone (PSU) matrices has proven especially effective for the removal of complex mixtures of contaminants. In particular, PSU-GO hollow fiber architectures have demonstrated outstanding adsorption performance toward organic micropollutants, including bisphenol A, ciprofloxacin, ofloxacin, and a range of endocrine-disrupting compounds, as well as metal ions (Pb, Cu, Cr(III)) and per- and polyfluoroalkyl substances (PFAS) [85,86]. These materials exhibit adsorption capacities that are comparable to or exceed those of granular activated carbon (GAC), the current industrial benchmark, achieving adsorption performance 3 to 7 times higher than GAC for heavy metals and short chain PFAS, while preserving mechanical robustness and operational stability.

Within this context, an innovative strategy involves the implementation of a closed-loop system for the recovery and reutilization of industrial membrane production waste [87]. Offcuts generated during the fabrication of polysulfone and PSU-GO hollow fiber membranes have been effectively upcycled into granular PSU and PSU-GO adsorbents specifically engineered for water treatment applications. These millimetric granules exhibit high adsorption efficiencies in tap water contaminated with lead, diclofenac, and perfluorooctanoic acid (PFOA). This closed-loop valorization route, which transforms manufacturing residues into high-value adsorbent materials, constitutes a circular economy approach designed to minimize plastic waste while concurrently enhancing water purification technologies. By means of such upcycling strategies, the life cycle of membrane materials is prolonged, overall energy demand is reduced, and waste management costs are diminished,

thereby promoting more sustainable industrial practices aligned with emerging environmental regulations and policy frameworks.

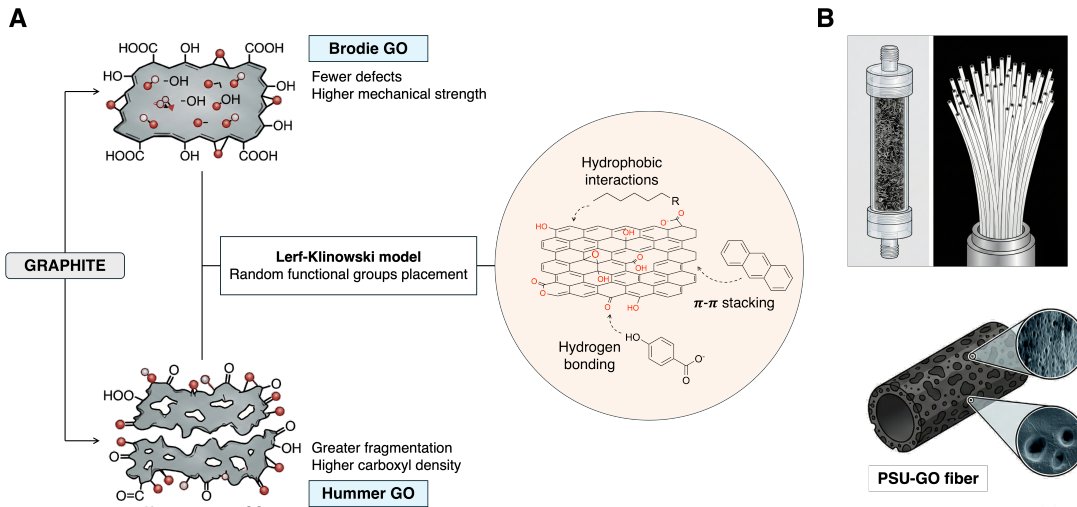


Figure 2: Graphene oxide structures and PSU-GO fibers. (A) Brodie and Hummer GO exhibit distinct structural properties, conventionally described by the Lerf-Klinowski model with random surface functionalization. (B) Schematic of a filtration cartridge utilizing pristine and upcycled porous PSU-GO hollow fibers, adapted from [85,87].

1.4 Aim of the thesis

The rapid proliferation of emerging contaminants poses a substantial risk to aquatic ecosystems and public health. Despite notable progress in the development of graphene-based and other nanostructured adsorbents, their large scale deployment is still hindered by the absence of a universal treatment technology capable of efficiently removing a chemically and structurally diverse array of pollutants. Because each contaminant class exhibits distinct interaction patterns with graphene oxide and other carbonaceous materials, the optimization of adsorbent design requires a detailed elucidation of the physicochemical driving forces that govern adsorption at the molecular scale.

In this context, computational modeling constitutes a powerful, yet comparatively underutilized, approach for expediting the rational design of next-generation adsorbent materials. Molecular simulation techniques, traditionally applied in computational biophysics to study ligand-protein recognition, afford atomistic resolution insight into the energetics, conformational landscapes, and thermodynamic signatures underlying adsorption phenomena. By adapting methodologies such as molecular docking and molecular dynamics (MD) simulations to problems in materials science, it

becomes feasible to characterize pollutant-adsorbent interactions *in silico*, elucidate the predominant binding mechanisms, and predict adsorption efficiencies prior to experimental validation. Such predictive frameworks have the potential to reduce the time, cost, and resource demands associated with conventional empirical screening campaigns.

To this end, the primary objective of this thesis is to develop and validate a multiscale, semi-automated computational screening pipeline to support the design of GO-based adsorbents for the removal of emerging contaminants from wastewater. The project seeks to integrate and optimize a suite of computational tools, including molecular docking, all-atom and coarse-grained MD simulations, into a coherent workflow capable of screening thousands of environmentally relevant pollutants. Through this strategy, the work aims to elucidate the molecular determinants of adsorption on GO surfaces, discriminate between weak and strong binders, and identify structural and chemical motifs that enhance adsorption performance.

Ultimately, the resulting computational platform is intended to serve as a predictive tool to guide the experimental synthesis, modification, and functionalization of graphene-based nanomaterials. By closely coupling molecular modeling with environmental nanotechnology, this research seeks to establish a new framework for virtual materials design, one that accelerates discovery, supports the development of sustainable remediation strategies, and provides a robust foundation for high-throughput assessment of pollutant-adsorbent interactions across diverse chemical classes.

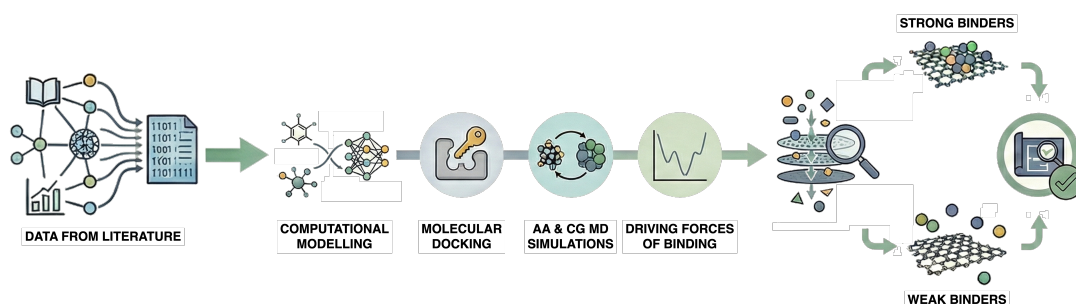


Figure 3: Aim of the thesis. The sequential pipeline leverages literature data, molecular docking, and MD simulations at multiple resolutions to understand the specific binding driving forces on graphene oxide models, supporting the rational, simulation-guided optimization of GO-based nanomaterials for wastewater treatment.

CHAPTER 2: DATABASE DESIGN

The present chapter describes the construction and characterization of two proprietary molecular databases. First, the emerging contaminants database (EC-DB) is introduced. This library comprises 435 unique parent compounds which, upon consideration of all associated protomers and stereoisomers, expand to approximately 2000 discrete molecular entities. These structures are classified into four principal categories: pesticides, pharmaceuticals, industrial chemicals, and priority pollutants.

Subsequently, the graphene oxide database (GO-DB) is presented. This library consists of four graphene oxide models exhibiting distinct degrees of oxidation, designed to represent the commercial LayerOne reference materials GOX8-P935, REGOX8-P980, REGOX8-P995, and REGOX8-P999. For all entries in both databases, force field parameters were generated using the AMBER GAFF2 force field, thereby enabling their integration into virtual screening pipelines and molecular dynamics simulations. The majority of the results described in this chapter were obtained using *in-house* Bash and Python scripts, which were developed to reduce manual intervention and enhance workflow reproducibility.

2.1 Methods

2.1.1 The PubChem Database

The initial phase of the project relied extensively on existing repositories to retrieve one dimensional (1D) molecular descriptors for each compound included in the pollutant library. In particular, the PubChem database was employed as the primary source of molecular descriptors and physicochemical properties.

PubChem is a publicly accessible repository of chemical substances and associated biological activity data, established in 2004 by the U.S. National Institutes of Health (NIH) [88]. It has since evolved into one of the largest open-access chemical databases worldwide, comprising more than 60 million unique chemical structures. It integrates data from a broad range of contributors, including academic laboratories, governmental organizations,

pharmaceutical companies, chemical suppliers, scientific publishers, other chemical databases (e.g., ChEMBL, DrugBank), and regulatory authorities such as the U.S. Food and Drug Administration (FDA) and the Environmental Protection Agency (EPA) [89].

A central feature of PubChem is its automated standardization workflow, which validates and normalizes all deposited chemical structures. For each standardized structure, PubChem automatically calculates a series of molecular properties (e.g., molecular weight, molecular formula, and partition coefficient logP) and generates chemical descriptors, including IUPAC names, canonical and isomeric SMILES, InChI, and InChIKey identifiers. Lastly, each chemically unique standardized structure is also assigned a unique Compound Identifier (CID). When a structure is processed, standardized, and determined to be novel within the database, a new CID is generated and associated with that compound [90].

In the context of the present retrieval pipeline, the following data fields are extracted from PubChem: CID, traditional and IUPAC names with associated synonyms, molecular formula, isomeric SMILES, InChIKey, and XlogP.

2.1.2 Molecule string notations

SMILES (Simplified Molecular Input Line Entry System) is a linear notation used to describe the structures of chemical species in the form of concise ASCII strings [91]. Developed in the late 1980s, SMILES remains the most widely adopted linear representation in cheminformatics because it enables computationally efficient storage, manipulation, and comparison of molecular structures.

In standard SMILES notation, atoms are represented by their elemental symbols. Elements belonging to the “organic subset” (B, C, N, O, P, S, F, Cl, Br, and I) may be written without square brackets when they possess the appropriate number of implicit hydrogen atoms. Atoms with explicitly specified hydrogen atoms, formal charges, or other non-standard valence states, as well as elements outside the organic subset, must be denoted within square brackets (e.g., [Au], [Fe], or [nH] for an aromatic nitrogen atom bearing an explicit hydrogen).

Covalent bonds are encoded using specific symbols: - for single bonds (which may be omitted by default), = for double bonds, # for triple bonds, and : for aromatic bonds. For instance, benzene is represented as c1ccccc1, where the lowercase *c* indicates aromatic carbon atoms, and the digit *1* marks the start and termination of a ring closure.

Isomeric SMILES extend the standard notation by incorporating stereochemical information, thus enabling the representation of defined stereoisomers of a given molecular framework. Two principal classes of stereochemistry are captured in isomeric SMILES:

Tetrahedral chirality

SMILES employs a local chirality convention based on the order in which atoms appear in the string, rather than on an absolute numbering system such as the Cahn-Ingold-Prelog (CIP) rules. The symbols @ and @@ indicate alternative configurations at stereogenic centers (commonly corresponding to clockwise and counterclockwise arrangements relative to the specified atom order), and they permit the partial or full specification of stereochemistry as required. For example, C[C@H](O)CC encodes (S)-butan-2-ol.

Geometric isomerism (double bond stereochemistry)

The configuration about double bonds is specified using directional bond symbols / and \, which denote the relative spatial orientation of substituents. For example, F/C=C/F or F\C=C\F denotes the trans (E) isomer of difluoroethene, where the fluorine atoms reside on opposite sides of the C=C bond, whereas F/C=C\F denotes the cis (Z) isomer, in which the fluorine atoms are located on the same side of the double bond.

Nonetheless, SMILES can encode an identical molecular structure in multiple distinct ways, leading to redundant, non-canonical representations. To address the lack of inherent canonicalization in SMILES, an alternative notation, the International Chemical Identifier (InChI), was developed. InChI is a standardized textual descriptor for chemical substances, officially released by IUPAC in 2005 [92]. In contrast to SMILES, each unique chemical structure corresponds to a single, unambiguous InChI string.

InChI is organized into a hierarchical, layered architecture comprising up to six distinct informational layers:

Main layer (M): contains the molecular formula, the atom connectivity graph (excluding bond order), and hydrogen connectivity. This is the only mandatory layer and must be present in every InChI;

Charge layer (Q/P): encodes formal charge and protonation state, which is critical for accurately representing ionizable compounds;

Stereochemical layer (B/T/M/S): describes stereochemistry, including double-bond geometry (sub-layer /b), tetrahedral stereocenters (sub-layer /t), and inverted stereochemistry (sub-layer /m);

Isotopic layer (I): Specifies the isotopic composition of atoms (e.g., enrichment in ^{13}C or ^2H);

Fixed-H layer (F): Represents hydrogen atoms that are fixed and not in equilibrium with the solvent, thereby capturing a specific tautomeric or protonation state;

Fixed/Isotopic combination layer (FI): Integrates the fixed-hydrogen and isotopic information to yield a fully specified structure.

Prior to InChI generation, input structures are subjected to a normalization step in which they are mapped onto a standardized parent structure. This procedure reduces the likelihood of generating distinct InChI identifiers for different tautomeric forms or closely related protonation states. For example, both acetic acid and its conjugate base, acetate, are normalized to the neutral acetic acid form. An illustrative InChI string is InChI=1S/C2H6O/c1-2-3/h3H,2H2,1H3, which encodes the molecular structure of ethanol.

As InChI gained widespread adoption in major chemical databases, it became evident that full InChI strings are often excessively long for efficient web searching and indexing. In response, the InChIKey was introduced in 2007 as a derivative identifier. The InChIKey is a condensed, fixed-length (27 character) digital representation of the corresponding full InChI, generated via the SHA-256 hashing algorithm. In contrast to InChI, the InChIKey is not designed to be human-readable; rather, it serves as a compact, standardized key for chemical structure identification. For instance, the InChI string for ethanol given above is converted to the InChIKey LFQSCWFLJHTTHZ-UHFFFAOYSA-N.

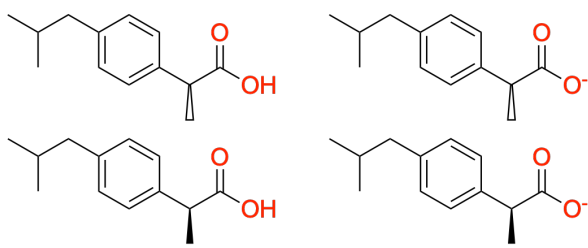
Because InChIKey generation is governed by a rigorously standardized algorithm, two independent chemical databases that contain the same chemical entity will derive identical InChIKeys, thereby facilitating robust cross-database integration and record matching [93]. This property renders the

InChIKey particularly advantageous for the rapid identification of identical structures and for the efficient retrieval of closely related compounds (e.g., those differing only in charge state or stereochemical specification) across heterogeneous chemical information resources.

Although it is not a molecular notation in the strict sense, a key molecular descriptor that is also provided by PubChem is XlogP, an estimate of the n-octanol-water partition coefficient. The partition coefficient (logP) characterizes the equilibrium distribution of a compound between n-octanol and water and is widely employed as a quantitative measure of lipophilicity or hydrophobicity: higher logP values correspond to increased lipophilicity, whereas lower (including negative) values reflect increased hydrophilicity.

The most widely adopted computational method for predicting logP is XLOGP3, which was developed via linear regression using a training set of approximately 8000 organic compounds with experimentally determined logP values [94]. XLOGP3 implements an additive atom-type approach: it predicts logP by summing the contributions of predefined atom types present in the molecule and subsequently refines this initial estimate by applying correction factors derived from comparisons with experimentally measured logP data.

PubChem Ibuprofen, (+-)- (Compound)



Property	Value
CID	3672
Formula	C ₁₃ H ₁₈ O ₂
IUPAC	2-[4-(2-methylpropyl)phenyl]propanoic acid
SMILES	CC(C)CC1=CC=C(C=C1)[C@H](C)C(=O)O
InChI	InChI=1S/C13H18O2/c1-9(2)8-11-4-6-12(7-5-11)10(3)13(14)15/h4-7,9-10H,8H2,1-3H3,(H,14,15)
InChIKey	HEFNNWSXXWATRW-UHFFFAOYSA-N

CONTENTS

- Title and Summary
- 1 Structures
- 2 Names and Identifiers
- 3 Chemical and Physical Properties
- 4 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Pharmacology and Biochemistry
- 9 Use and Manufacturing
- 10 Identification
- 11 Safety and Hazards
- 12 Toxicity
- 13 Associated Disorders and Diseases
- 14 Literature
- 15 Patents
- 16 Interactions and Pathways
- 17 Biological Test Results
- 18 Taxonomy
- 19 Classification
- 20 Information Sources

Figure 4: Structural identifiers of ibuprofen. PubChem database entry displaying the 2D molecular structures of ibuprofen enantiomers, alongside essential chemical identifiers such as its IUPAC name, SMILES, and InChI strings used for compound characterization.

2.1.3 Quantum Mechanics (QM) and AMBER parametrization

The development of both structural databases also required the application of quantum mechanical (QM) methods for parametrization, thereby ensuring that each molecule is suitable for use in molecular dynamics simulations. Quantum mechanics describes subatomic entities, including electrons, in terms of both particle-like and wave-like behavior, which leads to intrinsic non-locality and, consequently, to the fundamental indistinguishability of individual electrons.

According to Heisenberg’s uncertainty principle, the position of an electron cannot be determined exactly; instead, the electron is characterized by a probability of being found at a given point in space. This probability distribution is described by the wave function, which constitutes the solution of the time-independent Schrödinger equation:

$$H\psi(x, y, z) = E\psi(x, y, z) \tag{1}$$

Where: H is the Hamiltonian operator, ψ the wave function (eigenfunction) and E the energy of the system. The Schrödinger equation can be solved exactly only for hydrogen-like atoms, that is, for one-electron (monoelectronic) systems. For systems of greater complexity, it is necessary to introduce approximations in order to render the problem solvable.

The first fundamental approximation is the Born-Oppenheimer (BO) approximation, which enables the separation (decoupling) of electronic and nuclear motion [95]. Since nuclei are much heavier than electrons, their kinetic energy is smaller by a factor of approximately 10^5 , and the electronic motion can thus be treated in the presence of an effectively fixed nuclear framework. Within this framework, the nuclear kinetic energy operator can be neglected, and the internuclear repulsion term can be regarded as a constant for any given nuclear geometry:

$$\Psi_{total} = \psi_{nuclear}\psi_{electronic} \tag{2}$$

This facilitates the partitioning of the Hamiltonian operator into distinct electronic and nuclear components, thereby allowing the resulting reduced and effectively decoupled subsystems to be treated independently:

$$H_e + T_n \tag{3}$$

$$H_e\psi_e = E_e\psi_e \quad (4)$$

Still, even upon neglecting the nuclear kinetic energy term, no closed-form analytic solutions are available for many-electron systems because of electronic correlation effects. The Hartree-Fock Self-Consistent Field (HF-SCF) method is an *ab initio*, iterative procedure that further simplifies the electronic-structure problem by treating electro-electron repulsion within a mean-field approximation, in which the effect of the remaining $n - 1$ electrons on a given electron is represented by an average potential. In contemporary quantum chemistry, a variety of post-HF methods have been developed that provide systematically improved approximations to the exact electronic wave function and energy.

In practice, molecular and atomic orbitals are expanded in a basis set using the linear combination of atomic orbitals (LCAO) approach [96]:

$$\Phi = \sum_{i=0}^N a_i \psi_i \quad (5)$$

Where Φ is the newly composed molecular orbital, and ψ and a are the atomic orbitals considered and their coefficients. Gaussian-type orbitals (GTOs) currently constitute the most widely employed basis sets in electronic structure calculations, owing to their substantial computational advantages over the original Slater-type orbitals (STOs).

A particularly convenient class of basis sets is the family of split-valence basis sets introduced by Pople [97], in which core and valence orbitals are treated differently. The conventional notation for these basis sets is X-YZG. In this notation, X denotes the number of primitive Gaussian functions contracted to form each core atomic orbital basis function. The symbols Y and Z indicate that each valence orbital is represented by two contracted basis functions, the first being a linear combination of Y primitive Gaussians and the second a linear combination of Z primitive Gaussians. The use of two integers after the hyphen therefore signifies that the basis set is a split-valence double-zeta basis set. Split-valence triple- and quadruple-zeta basis sets are also employed, and are denoted as X-YZWG, X-YZWVG, and so forth. Pople-type basis sets may additionally incorporate polarization functions (e.g., 3-21G*), which increase the flexibility of the basis set in describing the anisotropy of the electron density [98]. A single asterisk indicates that polarization functions are included

only for atoms beyond the first period, whereas a double asterisk signifies that polarization functions are also added to hydrogen atoms. Furthermore, to improve the description of the diffuse tail region of atomic orbitals far from the nucleus, diffuse functions are often appended to the basis set; these are denoted by a + symbol (e.g., 3-21+G*) [99].

Since MD simulations, discussed in greater detail in Chapter 3.1.2, are performed using the AMBER25 package [100–102] in conjunction with the GAFF2 force field [103], the parametrization of the system was carried out following the standard three-step AMBER protocol [104], as outlined below. Any deviation from this established procedure would lead to inconsistencies in the derived partial charges relative to those of other molecular entities (such as amino acids, water, etc.) already parametrized within the AMBER/GAFF framework [105], thereby undermining the transferability and internal consistency of the overall model.

1. QM geometry optimization at the HF/6-31G* level

A QM geometry optimization is carried out *in vacuo* at the Hartree-Fock level with the 6-31G* basis set using Gaussian 16 [106]. This procedure yields the optimized, minimum-energy three-dimensional conformation that is subsequently employed as the reference structure for all downstream parametrization steps. Notably, the HF/6-31G* wavefunction was adopted historically not because it provides superior intrinsic accuracy, but because it systematically overestimates the gas-phase dipole moment by approximately 20%. This systematic overestimation compensates for the fact that RESP charges derived from gas-phase calculations are intended for use in condensed-phase (solvated) systems during MD simulations [107]. In the absence of this overestimation, the resulting partial charges would tend to underestimate electrostatic interactions in aqueous or protein environments, which in turn would degrade the reliability of predicted binding affinities and other molecular properties.

2. Restrained Electrostatic Potential (RESP) charge derivation

Based on the optimized geometry, partial atomic charges are derived using the RESP (Restrained Electrostatic Potential) method applied to the HF/6-31G* electronic density [105]. RESP is a regularized charge-fitting approach in which atomic charges are determined by fitting to the Merz-Kollman (MK) electrostatic potential (ESP) [108]. Instead of assigning charges empirically, RESP minimizes the deviation between the QM-calculated ESP and the ESP

generated by a set of atom-centered point charges. In addition, RESP incorporates symmetry and chemical equivalence constraints: for example, hydrogen atoms bound to the same carbon are constrained to have identical charges, reflecting their equivalent local chemical environments.

3. Atom-type assignment and force-field parameter generation

In the final stage, atom types and the corresponding force-field parameters are assigned automatically using the Antechamber [104] module distributed with AmberTools [102]. Antechamber infers bond connectivity and atom types from the local chemical environment, including hybridization state, ring membership, and functional group identity, and it groups atoms that are chemically analogous. Furthermore, it detects any missing force-field parameters (bond, angle, dihedral, and van der Waals terms) that must be supplied manually. Upon successful completion of the parametrization workflow, the following output files are produced: (i) an *frcmod* file, which contains any bond, angle, and dihedral parameters that supplement or extend the standard GAFF2 library; (ii) a *prepc* file, which encodes the molecular geometry, atom types, and partial charges; (iii) *parm7* and *rst7* files, which are AMBER topology and coordinate files, respectively, suitable for direct use in MD simulations; and (iv) a PDB file corresponding to the QM-optimized structure.

2.2 Emerging contaminants database (EC-DB)

Initially, a list of compound names was retrieved from Yang et al. [109] and categorized into three classes: pesticides, pharmaceuticals, and industrial compounds. A fourth class, hereafter referred to as priority compounds, was also defined. This group comprises selected molecules drawn from all three aforementioned categories and was established on the basis of prioritization criteria reported in the literature. The original list was used without modification and therefore contains redundant entries and occasional grammatical inconsistencies in the compound names.

The name list was subsequently processed using the following *in-house* workflow. For each compound, its name was screened and corrected for grammatical errors, then queried in the PubChem database via the *pubchempy* interface. The corresponding CID, traditional and IUPAC names (including synonyms), molecular formula, isomeric SMILES, and InChIKey were retrieved and stored. These metadata were then recursively cross-compared across all

entries in this initial database in order to remove redundancies present in the original list.

The isomeric SMILES were then processed with RDKit [110] to generate an initial three-dimensional (3D) structure for each compound (origin.pdb). Stereoisomerism and protonation states were addressed by two parallel procedures. Stereoisomer enumeration was performed using the *EnumerateStereoisomers* function implemented in RDKit, employing the SMILES string and the parameter set reported in Table 1. The parameter values were selected to mitigate combinatorial explosion and to prevent the generation of non-physical structures. Combined with the intrinsic structural curation and classification implemented in PubChem, this strategy ensures that the resulting enumerated molecules correspond to commercially available species and, for example, allows discrimination between racemic drug formulations and the specific pharmacologically active stereoisomer.

Table 1: Parameters for isomer generation. Summary of RDKit settings used to enumerate stereocenters and generate 3D coordinates, ensuring the exclusion of duplicates and sterically impossible isomers.

Parameter	Value	Description
onlyUnassigned	True	Enumerate stereochemistry at centers where it is not already defined
maxIsomers	32	Maximum number of isomers to yield
unique	True	Prevents duplicates by keeping stereoisomers with different SMILES
tryEmbedding	True	Discard sterically impossible isomers by attempting 3D coordinates generatio

While stereoisomer generation is fully automated, the protonation step is only partially automated and still requires manual intervention. The appropriate protonation state is selected by comparing experimentally reported pKa values with those predicted using the Chemaxon [111] software suite. In line with the sustainability-oriented scope of this work, all protomeric states relevant within the pH range 5-9, typical of environmental studies [112–117], were considered.

The initial structure, origin.pdb, is used as input and manually edited with GaussView 6 [118] and Avogadro 2 [119] to generate the desired protonation state. The resulting protonation-adjusted molecule is stored in a standardized format:

molecule_charge_protomer_template.pdb

where *molecule* denotes the compound name (e.g., aspirin, acetylsalicylic-acid), *charge* is encoded as a letter plus the net charge (e.g., N = 0, M1 = -1, P1 = +1, P2 = +2), and *protomer* is an integer index (1, 2, ...) distinguishing different protomers with the same net charge. For example, aspirin_M1_1_template.pdb corresponds to the first negatively charged protomer of aspirin. As implied by this nomenclature, the generated protomer template does not contain stereochemical information and is used solely as a protonation reference.

Protonation is propagated to all stereoisomers via an automated, template-driven hydrogen transfer protocol. The template molecule (with the selected protonation state) and the target stereoisomer are simultaneously loaded and spatially aligned using only heavy atoms, thereby ensuring structural correspondence independent of hydrogen positions. Hydrogen connectivity is then mapped by associating each hydrogen atom with its bound heavy atom, and discrepancies between the template and target are identified in terms of excess or missing hydrogens at each heavy-atom site. Surplus hydrogens in the target are removed, whereas missing hydrogens are explicitly added by copying the corresponding atoms and bonds from the template. The initial coordinates of the added hydrogens are inherited from the aligned template and translated by 1 Å to mitigate steric clashes.

The resulting protonated three-dimensional structures are subsequently subjected to geometric optimization using Open Babel’s Universal Force Field (UFF) [120,121] energy minimization (500 steps; convergence threshold of 0.01 kcal/mol). A post-processing step then removes any hydrogens automatically introduced during optimization owing to Open Babel’s valence rules. This workflow ensures that the final structures preserve the protonation pattern of the template while maintaining the stereochemical configuration of each target stereoisomer.

The final naming scheme for each processed structure is:

config_molecule_charge_protomer

where *config* encodes the E/Z, R/S, or U (unassigned) stereochemical descriptors. Once the curated set of correctly protonated 3D structures has been generated, each compound is parametrized according to the standard AMBER protocol, as detailed in the Methods 2.1.3 section. The database is

then finalized by organizing the directory structure for each molecule and removing superfluous temporary files.

The resulting collection of parent pollutants comprises 435 unique compounds, categorized into 74 pesticides, 141 industrial chemicals, and 220 personal care products. Within this total, a curated subset of 53 molecules has been designated as priority compounds. Accounting for all stereoisomers and protomeric forms of these 435 parent structures, the complete EC-DB database expands to contain approximately 2000 distinct molecular entries. A schematic overview of the EC-DB construction workflow is provided in Figure 5.

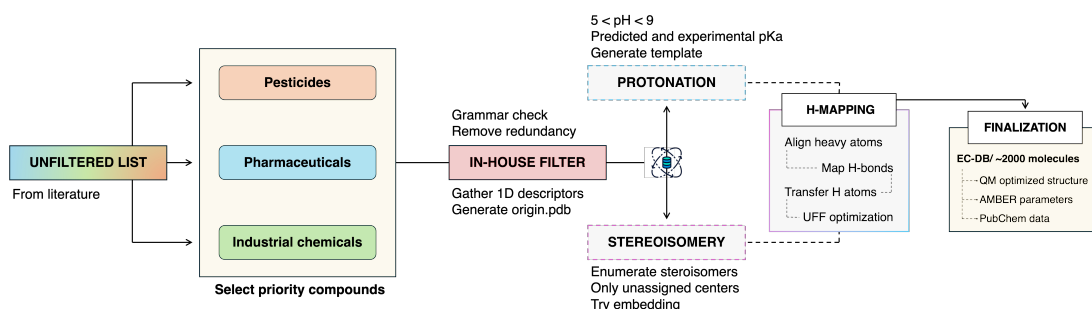


Figure 5: Workflow for the development of the Emerging Contaminants Database (EC-DB). An unfiltered list of priority compounds (pesticides, pharmaceuticals, and industrial chemicals) undergoes *in-house* filtering, protonation state assignment, and stereoisomer enumeration. Subsequent structural optimization and finalization yield a curated database of approximately 2000 molecules with optimized parameters.

2.3 Graphene oxide database (GO-DB)

To ensure the industrial relevance and reproducibility of the EC-DB, graphene oxide structures were constructed in relation to four commercially available LayerOne products (LayerOne AS, Oslo, Norway): GOX8-P935 [122], REGOX8-P980 [123], REGOX8-P995 [124], and REGOX8-P999 [125]. These commercial materials were used as reference benchmarks to define four distinct oxidation regimes, ranging from highly oxidized to nearly pristine graphene.

The O/C ratios of the resulting computational models (0.43, 0.19, 0.07, and 0.02) display slight deviations from the bulk technical specifications reported by the manufacturer. Such discrepancies are intrinsic to atomistic modeling, in which stoichiometry is constrained by a finite number of atoms and the specific local arrangement of functional groups, rather than reflecting the macroscopic, ensemble-averaged composition characteristic of bulk powders. Despite these

local deviations, the models faithfully reproduce the relative ordering of O/C ratios and the qualitative surface chemistry trends of the reference products, thereby providing a structurally consistent library suitable for investigating adsorption mechanisms across a broad range of pollutants.

In contrast to small molecules, GO is not parameterized as a single rigid entity but is instead constructed by combining multiple functional groups (FGs), in line with the widely adopted Lerf-Klinowski structural model [53], in which different oxygen-containing groups are randomly distributed on the graphene basal plane and edges. Representative GO functional groups, namely epoxides, carboxyls, hydroxyls, and carbonyls, were parameterized following standard AMBER protocols using a set of graphene nanoflakes functionalized with (i) an individual, unique FG and (ii) multiple FGs simultaneously, in order to capture charge delocalization effects.

Partial charges for in-plane sp^2 carbon atoms not directly bonded to any functional group were set to zero, consistent with literature reports indicating that the majority of the charge density is localized at the FGs [126–128]. Analogously to the generation of *prepc* files, a dedicated *lib* file specifying atom types and partial charges was created for each individual FG. The corresponding *frcmod* file describing the complete GO tile was then produced by manually refining the initial parameters suggested by the *parmchk2* utility within AmberTools24 [102].

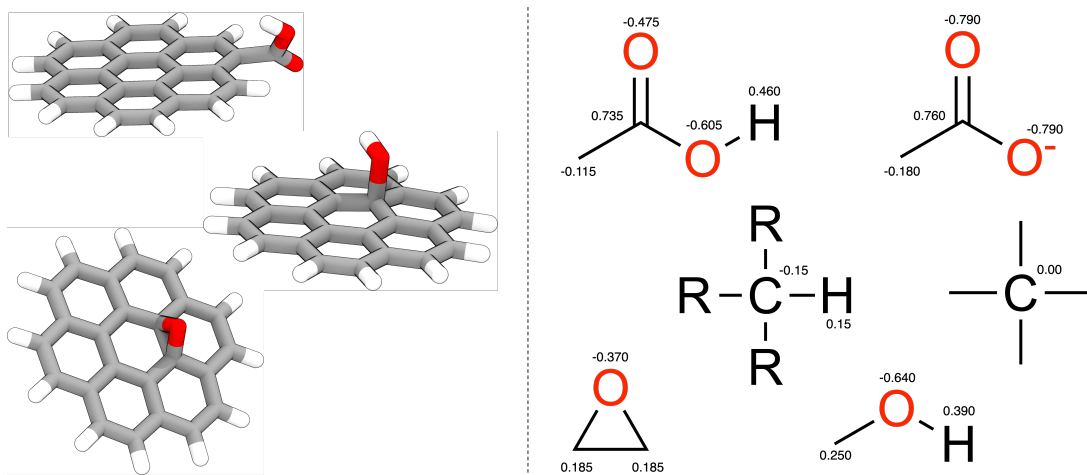


Figure 6: Parametrization of graphene oxide functional groups. 3D representations of graphene-like flakes (**left**) featuring common graphene oxide moieties: carboxyl, hydroxyl, and epoxide groups. The corresponding 2D representation (**right**) detail the assigned partial atomic charges with the RESP method.

The construction of GO surfaces is performed through a sequential protocol: (1) generation of a pristine graphene lattice; (2) introduction of vacancy defects associated with the oxidation process; (3) surface functionalization with primary GO functional groups, namely epoxide, carboxyl, and hydroxyl moieties; (4) deprotonation of carboxyl groups to yield carboxylates; and (5) addition of missing terminal CH hydrogen atoms, with the optional inclusion of carbonyl groups.

Steps 1-3 employ a combination of HierGO [129], a modular hierarchical software package for the generation of complex GO sheets originally parameterized for the OPLS-AA force field, and GOPY.py [130], an open-source tool for the construction of simple graphene oxide models. The HierGO source code was modified and adapted to incorporate the atom types and parameters of the GAFF2 force field, whereas the *lib* files distributed with the original GOPY.py repository were used as a starting point for the AMBER-compatible parametrization of each functional group. Steps 4 and 5 were implemented via *in-house* Python scripts.

Initially, a graphene sheet is generated by specifying the percentage of vacancy defects, where 0% corresponds to an ideal, defect-free graphene surface. The chosen vacancy fraction reflects the oxidation state of the target GO system, with highly oxidized materials exhibiting a more defective carbon framework [65,131,132]. Vacancies are introduced at random positions using a mixed-size distribution (small, medium, and large) drawn from the HierGO hole library. The resulting graphene surface is subsequently cleaned by removing dangling carbon atoms and then functionalized using GOPY.py through the random placement of epoxide, carboxyl, and hydroxyl groups. The total number and relative proportions of these functional groups are selected to reproduce the X-ray photoelectron spectroscopy (XPS) data of commercially available GO samples characterized by collaborators at ISOF-CNR [133,134].

Protonation states are then assigned by randomly converting 90% of the carboxyl groups to their corresponding carboxylate forms, reflecting the acidity and experimentally determined pKa values of GO edge functionalities in typical environmental and neutral pH conditions [135,136]. The Python workflow additionally supports manual control of deprotonation by allowing the user to specify individual carboxyl residues. Atom types are subsequently assigned to the preliminary GO structure, and the distinct fragments are assembled by forming covalent bonds between the different moieties via the

tleap *bondbydistance* command. Missing atoms, such as terminal CH hydrogens, are automatically added by tleap according to the definitions contained in the *lib* file templates.

To verify the numerical stability of the constructed models, a short (1 ns) *in vacuo* molecular dynamics simulation is carried out. Carbonyl groups can be optionally introduced at this stage by substituting selected terminal CH hydrogens. Using this procedure, graphene oxide sheets of effectively infinite lateral dimensions and with diverse surface chemistries can be generated, making the approach suitable for a wide range of applications, including the study of interactions with small molecules and macromolecular complexes, as well as the optimization of GO-based adsorbent materials. A schematic overview of the entire GO-DB generation workflow is presented in Figure 7, together with the four LayerOne graphene oxide models.

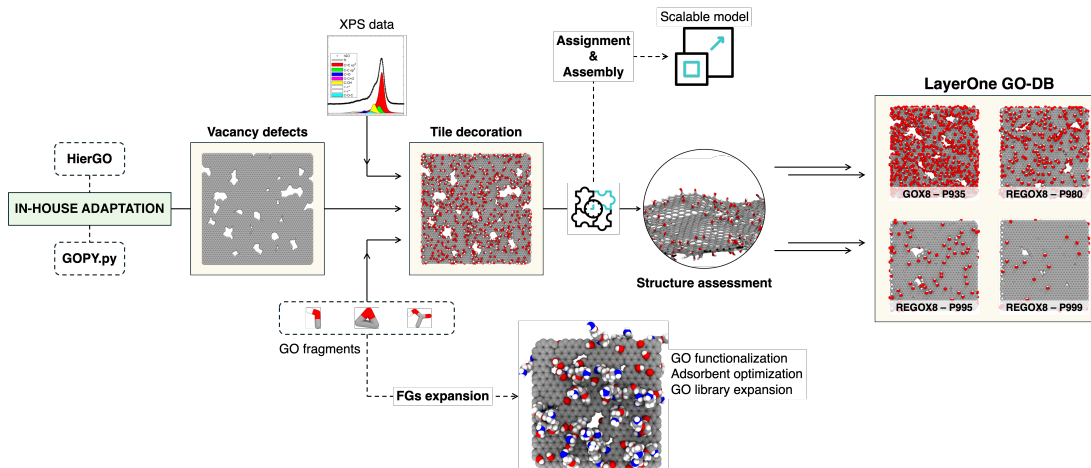


Figure 7: Workflow for the development of the LayerOne Graphene Oxide Database (GO-DB). The workflow integrates experimental XPS data to guide the distribution of oxygenated functional groups (tile decoration) across defective carbon lattices. The final output is the LayerOne GO-DB, tailored for computational screening and molecular dynamics simulations.

CHAPTER 3: VIRTUAL SCREENING

The two curated databases were integrated into the virtual screening (VS) workflow described in this chapter. In view of the expected computational cost, the screening was confined to priority compounds at pH 7, and the GOX8-P935 and REGOX8-P980 models were selected owing to their predominant use in previous experimental studies. The workflow consists of two principal, sequential stages: (1) grid-based docking followed by all-atom molecular dynamics refinement, and (2) prediction of the binding affinities of the resulting EC@GO complexes.

Because of the intrinsic stochastic character of the docking step, the VS protocol was systematically optimized to balance computational efficiency, predictive accuracy, and statistical robustness. Considering all combinations of the 53 priority pollutants, their stereoisomers, and the two graphene oxide models, the screening yielded approximately 2900 EC@GO complexes, corresponding to an aggregate simulation time of roughly 29 microseconds.

The results demonstrate that adsorption is predominantly determined by the relative orientation of the ligand on the chemically heterogeneous GO surface, rather than by its stereochemical configuration, and that inclusion of entropic contributions is essential to correct the size dependence inherent to enthalpy-only scoring functions. The final thermodynamic ranking reveals a prevalence of hydrophobic and π - π stacking interactions; polycyclic aromatic hydrocarbons and plasticizers display the highest affinities, whereas polar and anionic species are disfavored due to substantial de-solvation penalties.

The resulting dataset provides both a rigorous methodological benchmark for computational high-throughput screening and a valuable decision-making framework for prioritizing subsequent experimental investigations.

3.1 Methods

3.1.1 Virtual screening and docking

The methodologies discussed in this chapter originate from biophysics and are routinely employed in drug discovery workflows. These approaches have

subsequently been adapted and implemented in materials science, and, in the present context, applied to graphene oxide. Virtual screening (VS) techniques emerged as an adaptive response to the advent of high-throughput synthesis and screening technologies and currently constitute a core strategy for exploring vast regions of chemical space in contemporary drug discovery approaches [137].

VS encompasses a set of computational procedures designed to interrogate large libraries of small molecules in order to identify which candidates most likely bind to a given pharmacological target, typically a protein receptor or enzyme. The process is generally organized as a sequential and hierarchical pipeline aimed at filtering and prioritizing compounds that display suitable physicochemical, pharmacokinetic, and pharmacodynamic properties.

Initially, a virtual library is curated and standardized into a single, machine-readable format; molecular conformers are generated, and stereochemical inconsistencies as well as valence errors are corrected. Subsequently, computational methods are employed to estimate binding affinities, after which selected candidates proceed to experimental validation through *in vitro* and *in vivo* assays. The final output is a ranked list of molecules, where higher-ranking compounds exhibit an increased likelihood of manifesting measurable biological activity and ultimately progressing as drug leads. Thus, starting from a chemical space comprising billions of potential structures, VS seeks to identify a tractable subset of candidate molecules with putative activity [137–139].

Two principal VS strategies can be distinguished, which may be applied independently or in an integrated manner. Ligand-based virtual screening (LBVS) does not require explicit knowledge of the receptor structure; instead, it exploits information encoded in known ligands. In this approach, the virtual library is filtered using similarity search algorithms, quantitative structure-activity relationship (QSAR) and related modeling strategies, machine-learning methods, or ligand-based pharmacophore models, to identify compounds that are structurally or pharmacophorically similar to established actives [138]. In contrast, structure-based virtual screening (SBVS) relies on the three-dimensional structure of the target receptor, typically employing molecular docking, *de novo* design methodologies, or structure-based pharmacophore modeling.

In both LBVS and SBVS workflows, molecular docking serves as a central reference technique and can be integrated into all stages of the drug design process, including lead optimization [140]. Specifically within structure-based VS, docking is systematically applied to all compounds in a database against a specified receptor structure, yielding a ranking that reflects the predicted binding affinity between each ligand and the target receptor [141].

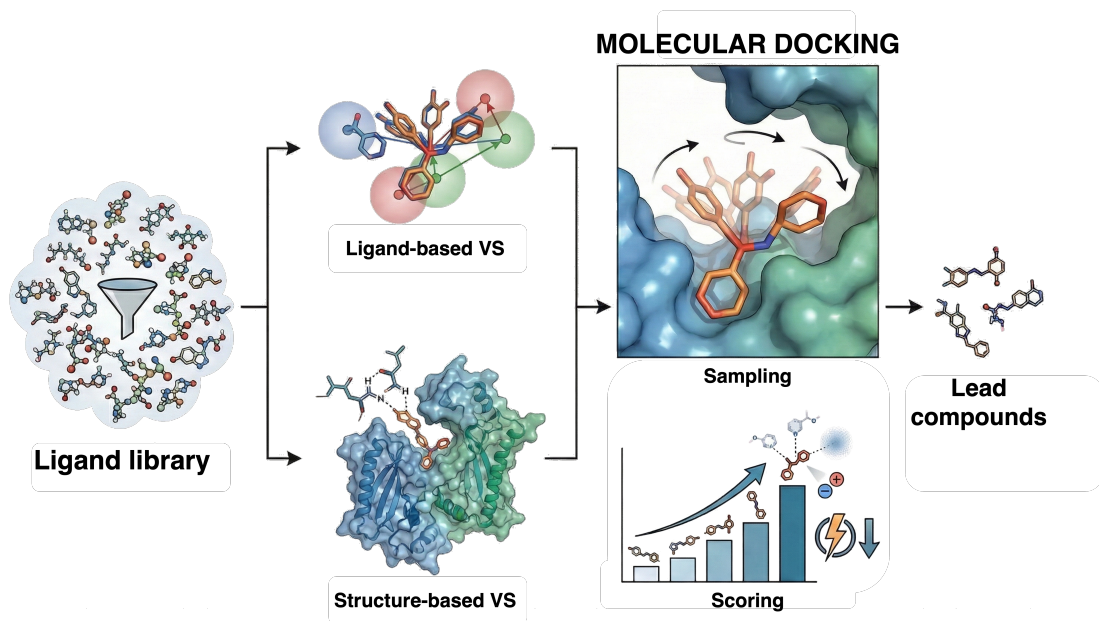


Figure 8: Molecular docking and virtual screening process. Starting from a diverse ligand library, candidates are prioritized through ligand- and structure-based VS techniques. Subsequent molecular docking assesses binding poses and interaction energies to isolate the most favorable lead compounds for further analysis.

Molecular docking is a structure-based drug design methodology that aims to characterize the geometric and energetic complementarity between a ligand and its macromolecular receptor. Conceptually, docking is often introduced via the lock and key model, which idealizes ligand-receptor recognition as the stereospecific insertion of a ligand into a preformed binding pocket on the receptor. Within this conceptual framework, the outcome of a docking calculation is a pose, defined as a particular geometry and orientation of the ligand relative to the receptor, yielding a specific configuration of the complex. In practice, however, the binding process is more accurately described by induced-fit or conformational selection models, in which both ligand and receptor can undergo conformational rearrangements upon complex formation. Molecular docking therefore represents an approximate computational reconstruction of this process, relying on simplifying assumptions to ensure tractability in terms of time and resources [140,142].

Docking methodologies are commonly categorized along several criteria [143]. A first classification distinguishes bound from unbound docking. In bound docking, an experimentally determined structure (typically X-ray crystallography or cryo-EM) of the receptor already exists in complex with a ligand. The receptor and ligand are computationally separated, and the docking procedure is applied to reproduce the experimentally observed pose. This protocol is primarily used for benchmarking and validation, as the known binding mode enables direct comparison between predicted and experimental poses to assess method accuracy [144,145]. In unbound docking, by contrast, the structure of the ligand-receptor complex is not known a priori. The objective is to predict the most favorable binding pose of a ligand within a receptor binding site without prior detailed knowledge of the exact interaction pattern.

A second, complementary classification differentiates local from global docking. Local docking assumes partial or complete prior knowledge of the binding site; the algorithm then explores alternative ligand positions, orientations, and sometimes conformations within this predefined region. Global docking, in contrast, does not presuppose any information about the binding site location and instead searches for potential interaction regions across the entire receptor surface [146,147].

Docking methodologies can be further classified according to their treatment of molecular flexibility [148,149]. In rigid-body docking, both ligand and receptor are considered inflexible entities; only the six external degrees of freedom of the ligand (three translational and three rotational) are sampled relative to the receptor [150]. While this approximation is computationally efficient, it deviates substantially from the physicochemical reality of molecular recognition and is therefore employed relatively infrequently. In contrast, flexible docking explicitly accounts for conformational degrees of freedom. A widely adopted pragmatic compromise, implemented in many contemporary docking programs, is to model the receptor as rigid while allowing the ligand to be flexible, typically by sampling its internal torsional degrees of freedom [148]. This approach captures a substantial fraction of the relevant conformational landscape at an acceptable computational cost. The most physically realistic framework would permit full flexibility of both ligand and receptor; however, the associated conformational search space becomes computationally prohibitive [151]. Consequently, receptor flexibility is in practice only partially incorporated, often restricted to selected regions (e.g.,

side chains in the binding site) and frequently applied only during later refinement stages [152]. As a result, discrepancies between docking-generated complexes and experimentally determined structures can be considerable, especially when the ligand is highly flexible and separated by relatively low energy barriers between alternative conformational states.

The docking workflow begins with preparation of the receptor and ligand structures. Typically, solvent molecules, co-crystallized ligands, and small ions not directly involved in binding are removed to expose possible binding pockets [153,154]. Once the structures are prepared, the docking protocol generally proceeds in two main steps: sampling and scoring.

During the sampling step, multiple docking poses are generated using search algorithms that explore the ligand conformational space and its placement relative to the receptor. From a computational standpoint, this step is relatively straightforward and can be implemented using systematic, stochastic, or hybrid search strategies [155]. In the subsequent scoring step, the poses generated by the sampling algorithms are evaluated and ranked. A scoring function assigns to each pose a numerical score that estimates the interaction energy or binding affinity of the ligand-receptor complex, with the aim of identifying the most plausible binding modes and the most promising ligands [156].

Despite their diversity, most scoring functions consider similar physical contributions. These include hydrogen bonding, geometric complementarity, polar and nonpolar contact terms, van der Waals interactions, electrostatic interactions, entropic contributions, and de-solvation effects [154,157]. The treatment of entropic and de-solvation terms is particularly challenging, as they are inherently complex to model and act as penalties for the binding process. Their accurate estimation is one of the main limiting factors in docking accuracy [156,158]. From a methodological standpoint, scoring functions are often classified into three main classes: empirical, force-field based, and knowledge-based [159].

Empirical scoring functions are derived from experimental measurements, such as quantified binding affinities, and model the binding free energy as a linear combination of physically motivated contributions with empirically fitted coefficients. A general empirical functional form can be expressed as follows:

$$\Delta G = \Delta G_0 + \Delta G_{hb}N_{hb} + \Delta G_{lipo}N_{lipo} + \Delta G_{bad}N_{bad} + \Delta G_{rot}N_{rot} + E \quad (6)$$

where the ΔG_i terms are energetic contributions associated with specific interaction types (e.g., hydrogen bonds, hydrophobic contacts, unfavorable contacts, rotatable bonds), the N_i terms are the corresponding counts or descriptors, and E is a residual term. The parameters are obtained by fitting to a training set of ligand-receptor complexes whose binding free energies are experimentally known or measurable.

Force-field-based scoring functions compute the noncovalent interaction energy between ligand and receptor using classical molecular mechanics expressions. In this case, the total nonbonded energy $E_{nonbond}$ is usually given by a sum of Lennard-Jones and Coulomb terms of the form

$$E_{nonbond} = \sum_{lig} \sum_{prot} \left[\frac{A_{ij}}{r_{ij}^6} - \frac{B_{ij}}{r_{ij}^{12}} + c \frac{q_i q_j}{r_{ij}} \right] \quad (7)$$

where the first two terms correspond to the van der Waals contribution modeled by the Lennard-Jones equation, and the last term represents the electrostatic interaction modeled by Coulomb’s law, with r_{ij} the interatomic distance, A_{ij} and B_{ij} Lennard-Jones parameters, q_i and q_j partial charges, and ϵ the dielectric constant.

Knowledge-based scoring functions are constructed from statistical analyses of structural databases of protein-ligand complexes. They estimate effective interaction potentials under the assumption that the observed frequencies of specific atom-atom contacts at defined interatomic distances are related to the underlying interaction free energies, typically via an inverse Boltzmann formalism.

Across different implementations, both the level of approximation used to represent each energetic contribution and the computational cost associated with their evaluation vary substantially, which in turn results in method-dependent performance across distinct docking scenarios. Consequently, molecular docking is not generally intended to directly identify clinically efficacious drug candidates but rather to filter large compound libraries, thereby reducing the number of molecules that must subsequently be characterized using more accurate, computationally demanding methodologies and, ultimately, experimental validation [139,154].

The primary docking engine employed in this thesis for structure-based virtual screening is AutoDock Vina [160,161]. AutoDock Vina is a widely adopted platform for flexible molecular docking, particularly advantageous when explicit treatment of ligand conformational flexibility is required. Under the standard protocol used here, the receptor is modeled as rigid, whereas the ligand is treated as flexible by specifying its rotatable bonds, such that internal torsional degrees of freedom are sampled during the docking process. AutoDock Vina utilizes a stochastic global optimization algorithm that iteratively adjusts the torsional angles and rigid-body position and orientation of the ligand relative to the receptor to identify low energy binding poses. The conformational search is confined to a user-defined region of space, specified as a three-dimensional box. These grid boxes are configured via the AutoDock Tools (ADT) interface, which allows the user to define both the center and the dimensions of the search volume [162].

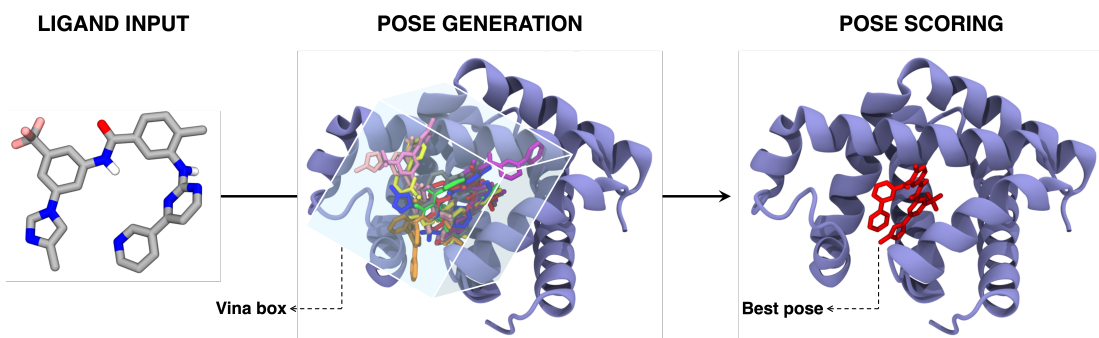


Figure 9: Docking process with AutoDock Vina. Various ligand conformers are sampled within a predefined search space (Vina box) and subsequently scored to isolate the best binding pose.

The scoring function implemented in AutoDock Vina is based on an empirical force-field formalism and provides an approximation of the binding free energy by combining steric, hydrophobic, hydrogen-bonding, and other noncovalent interaction terms. For each generated docking pose, the program reports an estimated binding affinity and an associated rank, and the output is presented as a list of poses sorted by increasing predicted binding energy (corresponding to decreasing binding affinity). The resulting protein-ligand complexes can be visualized and subjected to more detailed analysis using molecular graphics software such as VMD [163], which enable in depth examination of binding modes, intermolecular interaction networks, and ligand conformations within the binding site.

3.1.2 Molecular dynamics

Molecular dynamics (MD) is a computational methodology employed to simulate the temporal evolution of a molecular system [164,165]. In its classical formulation, atoms are typically modeled as rigid spheres and chemical bonds as harmonic oscillators characterized by suitable potential energy functions [166]. By numerically integrating Newton’s equations of motion, it is possible to obtain the trajectories of all atoms in the system and thereby quantify their displacements as a function of time.

MD represents a specific branch of molecular mechanics (MM) and is extensively used because it provides detailed information on the dynamical behavior of molecular systems over time. However, as a method derived from MM, a fundamental limitation of MD is the absence of an explicit electronic description. Consequently, processes involving bond breaking and formation cannot be captured within this purely classical framework. In this context, atoms are treated as point charges with excluded volume determined by their van der Waals radii, and bonds are commonly represented as ideal springs.

Overall, molecular dynamics simulation is a classical approach grounded in Newton’s second law of motion, which forms the basis for calculating atomic trajectories and exploring the time-dependent properties of molecular systems.

$$F = ma \tag{8}$$

Where F is the force exerted on the particle, m is its mass and a its acceleration. In order to obtain the latter for each atom of the system, it is necessary to know the force on each particle. Thus, the Newton’s motion equation is solved for each atom (i):

$$F_i = m_i a_i = m_i \frac{d^2 r_i}{dt^2} \tag{9}$$

Where r_i is the position of each atom. The force exerted can be expressed in function of the potential energy (V) as well, as in the following equation:

$$F_i = -\frac{\partial V}{\partial r_i} \frac{d^2 r_i}{dt^2} \tag{10}$$

This implies that, given the coordinates of an initial structure and an assigned set of velocities, the force acting on each atom can be computed provided that the potential energy function of the system is known.

The initial velocity distribution is typically generated by a stochastic procedure employing a Gaussian (Maxwell-Boltzmann) distribution consistent with the desired initial temperature. In contrast, the initial atomic positions are generally obtained from experimental structural data, most commonly derived from X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy. Once the forces, and thus the accelerations of all particles, have been evaluated using Eq. 9 and Eq. 10, the positions of the particles can be updated accordingly. The time evolution of the system, i.e., the trajectory, is then determined by expanding the positions in a Taylor series:

$$r(t + \delta t) = r(t) + v\delta t + \frac{a(t)}{2}(\delta t)^2 + \frac{b(t)}{6}(\delta t)^3 + \dots \quad (11)$$

$$v(t + \delta t) = v(t) + a(t)\delta t + \frac{a(t)}{2}(\delta t)^2 + \frac{b(t)}{6}(\delta t)^3 + \dots \quad (12)$$

Where $a(t)$ is the acceleration and $b(t) = a(t)$. The equivalent expansion of $r(t - \delta t)$ is:

$$r(t - \delta t) = r(t) - v\delta t + \frac{a(t)}{2}(\delta t)^2 - \frac{b(t)}{6}(\delta t)^3 + \dots \quad (13)$$

By adding the Eq. 11 to the Eq. 13 and substituting the acceleration with $\frac{F}{m}$, the following expression is obtained:

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + \frac{F(t)}{m}(\delta t)^2 \quad (14)$$

$$v(t) = \frac{r(t + \delta t) - r(t - \delta t)}{2\delta t} \quad (15)$$

The last two equations form the Verlet algorithm [167]. In principle, this procedure a long trajectory can be obtain with a timestep of δt . Nevertheless, at each step a numerical error is added, which should be kept under control. The Verlet integration scheme is introduced to reduce the numerical error associated with the computation based on Eq. 11. When the Taylor series is truncated after the second order term, the resulting local truncation error is of

third order in the time step. By summing Eq. 11 and Eq. 13, the leading error term is shifted to fourth order, thereby yielding a more accurate integration. Iterative application of this procedure generates a trajectory that represents the time evolution of the system.

The accuracy of the simulation is directly determined by the choice of the potential energy function V employed to describe the interactions between particles. In molecular dynamics simulations, a classical potential energy function (force field, FF) is commonly employed [166]. This function is defined in terms of the atomic coordinates of all particles comprising the system. The potential energy function arises from the sum of several terms representing covalent interaction, such as V_{bond} , V_{angle} , $V_{dihedral}$, and $V_{improper}$, and non-covalent interactions, such as $V_{Coulomb}$ and V_{LJ} [168]. The last correspond to interaction between particles separated by more than three covalent bonds. Thus, the potential function is the sum of the following terms.

For the bond lengths r_i , the bond angle θ_i , and the improper dihedral (out of plane) angle ξ_i , which represents the pyramiding of the i -th atom, the covalent interaction terms are represented by harmonic potential. For the dihedral angle ψ_i a more complex potential form is given:

$$V_{bond} = \frac{1}{2} \sum_{i=1}^{N_b} k_i^b (r_i - r_{0,i})^2 \quad (16)$$

$$V_{angle} = \frac{1}{2} \sum_{i=1}^{N_\theta} k_i^\theta (\theta_i - \theta_{0,i})^2 \quad (17)$$

$$V_{improper} = \frac{1}{2} \sum_{i=1}^{N_\xi} k_i^\xi (\xi_i - \xi_{0,i})^2 \quad (18)$$

$$V_{dihedral} = \frac{1}{2} \sum_{i=1}^{N_\phi} k_i^\phi (1 + \cos(\phi_i - \phi_{0,i})) \quad (19)$$

The non-covalent interactions are described by Coulomb's law, where q_i and q_j are the partial charge on the atom i and j , while the van der Waals interactions by a Lennard-Jones potential:

$$V_{Coulomb} = \sum_{i < j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) \quad (20)$$

Where A_{ij} and B_{ij} are the strength parameters for repulsion and attraction respectively. In the formulation of the Coulomb term, it is commonly assumed that each atom carries a partial charge. However, in an actual molecule, the electronic charge distribution is considerably more complex. The positive charge can be treated as point-like, as it is localized within the atomic nucleus, whereas the negative charge associated with the electrons is delocalized over the molecular volume. Consequently, the representation of the molecular charge distribution as a set of partial charges centered on atoms constitutes an approximation. Despite this simplification, such model is highly practical and is widely adopted in many empirical force fields [105].

Indeed, electrostatic interactions play a crucial role in determining both the conformational preferences of individual molecules and the nature of intermolecular interactions. It is therefore essential that partial charges be assigned with the highest possible accuracy. Multiple methodologies have been developed for the derivation of partial atomic charges, including, for example, Mulliken population analysis [169] and the RESP fitting procedure [105].

Lastly, force fields may differ not only in their functional forms but also in their parametrization strategies, as any given parameter set is typically optimized for a specific class of systems and cannot be transferred without loss of accuracy to arbitrary molecules. For instance, the AMBER family of force fields is primarily parameterized for biomolecular systems such as proteins, lipids, and nucleic acids [107]. As a result, nonstandard residues usually require dedicated, ad hoc parametrization [103].

A generic MD simulation protocol can be partitioned into three principal stages: (1) preparation of the parameter and topology files, (2) execution of energy minimization followed by equilibration, and (3) generation of the production dynamical trajectory. The energy minimization step is based on the evaluation of the gradient of the potential energy with respect to each atomic coordinate, which can be expressed as follows:

$$\frac{\partial f}{\partial x} = 0 \rightarrow \left(\frac{\partial V}{\partial \mathbf{r}} \right)_{i=1}^N = 0 \quad (21)$$

The calculation can be done through different methods [170]. First-order methods (steepest descents method, conjugated gradient minimization), in which the gradient is calculated and iterated, modifying the geometry by a

certain Δx in the direction of the gradient, until the gradient becomes null. However, a limitation of these methods is that they do not distinguish between maxima, minima, and saddle points of the objective function. This shortcoming can be addressed by second-order methods (e.g., the Newton-Raphson algorithm), in which the gradient is used in conjunction with the Hessian matrix to rigorously characterize the critical point and, under appropriate conditions (e.g., positive definiteness of the Hessian), to guarantee that it corresponds to a local minimum of the function.

The state (microstate) of a classical system can be completely characterized by specifying the positions and momenta of all the particles that constitute it. Such system can be represented by an ensemble, defined as an ideal collection consisting of a large number of replicas of the same system considered simultaneously, each replica corresponding to a possible microstate of the actual physical system. The members of the ensemble are governed by the same set of microscopic interactions and share a common set of macroscopic properties. Each replica evolves in time according to the underlying microscopic equations of motion, starting from different initial conditions, such that at any given instant the system occupies a well-defined, unique microstate.

Statistical mechanical ensembles are defined by boundary conditions [\[171\]](#):

In the canonical ensemble (NVT), the number of particles (N), volume (V), and temperature (T) are held constant;

In the microcanonical ensemble (NVE), the number of particles (N), volume (V), and total energy (E) are conserved;

In the grand canonical ensemble (μ VT), the chemical potential (μ), volume (V), and temperature (T) are fixed, while the number of particles may fluctuate;

In the isothermal-isobaric ensemble (NPT), the number of particles (N), pressure (P), and temperature (T) are maintained constant.

For large biomolecular systems, MD simulations become computationally intensive due to the high dimensionality and the long-range nature of intermolecular interactions. To alleviate this, MD software packages (such as AMBER) employ several methodological approximations and algorithmic optimizations, including periodic boundary conditions, implicit or simplified solvent representations, and interaction cutoffs, to enhance computational efficiency while minimally perturbing the potential energy surface.

Periodic boundary conditions (PBC) are commonly used to approximate bulk environments in MD simulations performed with explicit solvent [172]. A direct representation of a truly macroscopic solvent environment would require an effectively infinite number of particles, which is computationally prohibitive. Instead, the system is embedded in a finite simulation box (often cubic or truncated octahedral/hexagonal) containing a manageable number of solvent molecules. This primary cell is replicated in all spatial directions to form a three-dimensional tessellation, so that each particle experiences conditions approximating to the extended bulk phase. When an atom exits the primary simulation cell, its periodic image simultaneously enters from the opposite face, preserving the particle density and mimicking an effectively infinite system.

However, the introduction of PBC creates an additional complication: each particle can, in principle, interact not only with all other particles in the primary cell, but also with all of their periodic images, leading to a formally infinite number of pairwise interactions. For short-range interactions, such as those described by the Lennard-Jones potential V_{LJ} , the interaction strength decays rapidly with r^{-6} . Consequently, interactions beyond a chosen cutoff radius can be neglected with limited loss of accuracy, significantly reducing the computational cost. In contrast, electrostatic interactions decay as r^{-1} and therefore remain significant over long distances. Truncating Coulombic interactions at a finite cutoff introduces substantial artifacts. To address this, Ewald summation and its variants were developed to efficiently and accurately compute long-range electrostatics under PBC conditions [173].

Lastly, a variety of explicit water models have been developed for MD simulations. One widely used model is the rigid water model TIP3P, in which the O-H bond lengths and H-O-H angle are constrained to fixed values [174]. By treating the water geometry as rigid (i.e., removing intramolecular vibrational degrees of freedom of the hydrogens), the fastest bond-stretching motions are eliminated. This permits the use of a larger integration timestep without compromising numerical stability, enabling the sampling of slower, large-scale conformational motions over longer effective simulation times at a comparable computational cost.

3.1.3 MM-GBSA/MM-PBSA methods

The Molecular Mechanics/Poisson Boltzmann Surface Area (MM/PBSA) and the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA)

methods calculate binding free energies for macromolecules by combining molecular mechanics calculations and continuum solvation models [175]. The analysis of the binding affinity by which a ligand binds to its receptor can be evaluated by thermodynamics, where the reactions are driven by the minimization of the Gibbs free energy G :

$$G = U + pV - TS = H - TS \quad (22)$$

Where U is the internal energy, p pressure, V the volume, T the temperature, S the entropy and H the enthalpy of the system. If the Gibbs free energy is negative, the binding process occurs spontaneously. The contributions to the free energy are made explicit:

$$G = E_{int} + E_{el} + E_{vdW} + E_{pol} + E_{np} - TS \quad (23)$$

Where the first three terms are standard MM energy terms. E_{int} represents the internal bonded interactions (bond, angle, and dihedral), E_{el} the electrostatic ones, and E_{vdW} the van der Waals interactions. G_{pol} and G_{np} are the polar and non-polar contributions to the solvation free energies. G_{pol} is typically obtained by solving the PB equation or by using the Generalized Born (GB) model (giving the MM/GBSA approach), whereas the non-polar term is estimated from a linear relation to the solvent accessible surface area (SASA). The ΔG is calculated as follow:

$$\Delta G = G_{complex} - G_{ligand} - G_{receptor} \quad (24)$$

The value of ΔG provides a measure of the strength of the binding. Given the inherent challenges associated with directly quantifying binding free energies at the molecular scale using experimental techniques, computational methodologies for estimating these thermodynamic parameters have become indispensable in biochemical research, facilitating the prediction of binding affinities across a wide spectrum of biomolecular complexes. Such calculations are also crucial in computer-aided drug design, as they support the identification and prioritization of novel candidate molecules in the early stages of drug discovery.

The MM-PBSA/GBSA framework, first introduced in 1998 by Srinivasan and co-workers [176], is a post-processing technique that exploits molecular dynamics trajectories of the unbound components and the bound complex.

Within this approach, the absolute binding free energy is obtained from the ensemble-averaged free energies of the complex, the ligand, and the receptor, as follows:

$$\Delta G = \langle G_{complex} \rangle - \langle G_{ligand} \rangle - \langle G_{receptor} \rangle \quad (25)$$

The average values reported in brackets are obtained by post-processing a set of representative frames extracted from the molecular dynamics’ trajectories. In the MD simulations, the system (complex, ligand, and receptor) is embedded in a periodic box containing explicit solvent molecules. Within the MM-PBSA/GBSA framework, this explicit representation is replaced by a continuum solvent model to reduce the overall computational cost.

In principle, the absolute binding free energy in solution ΔG_{bind}^{solv} could be computed directly in the fully solvated state. However, such calculation would yield an unreliable estimate because the dominant contribution to the total energy would arise from solvent-solvent interactions. To mitigate these large energetic fluctuations and to isolate the relevant contributions, the MM-PBSA/GBSA methodology employs a thermodynamic cycle, as depicted in [Figure 10](#).

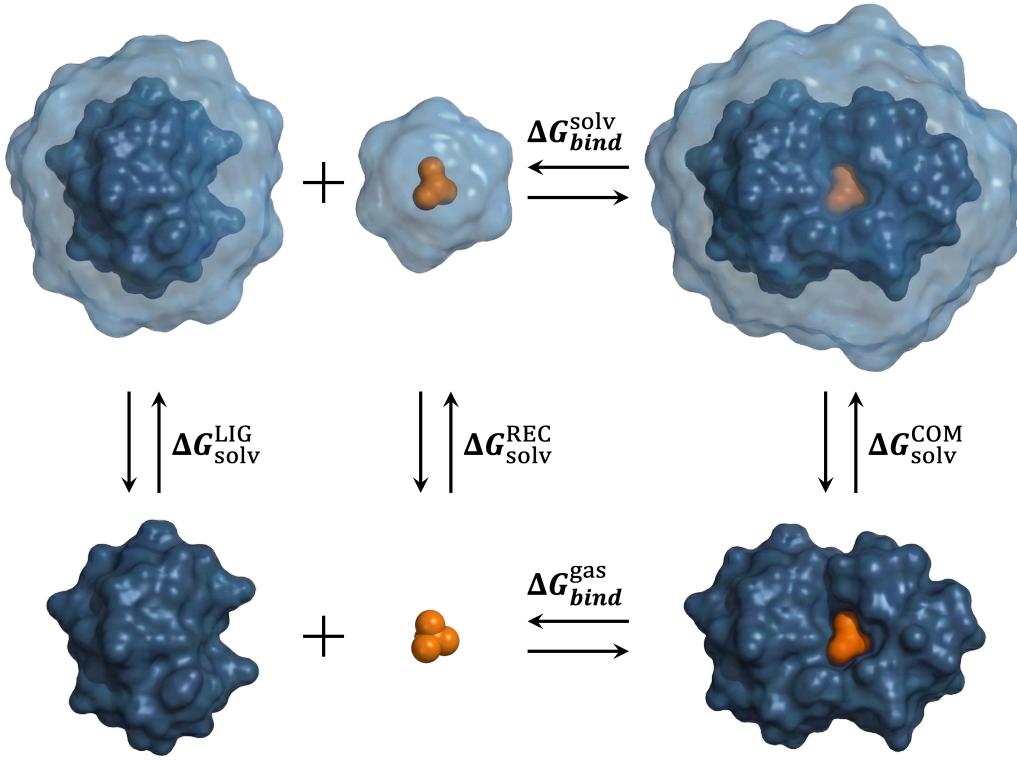


Figure 10: MM-GBSA thermodynamic cycle. Representation of the continuum solvent approach used to estimate binding affinities. The method calculates the final solvated binding free energy by assessing the interactions in the gas phase and applying corrections for the de-solvation penalties of the free receptor and ligand upon complex formation.

The binding free energy ΔG_{bind}^{solv} in solution is calculated as:

$$\Delta G_{bind}^{solv} = \Delta G_{bind}^{gas} - \Delta G_{solv} \quad (26)$$

$$\Delta G_{bind}^{gas} = [\Delta G_{gas}^{COM} - (\Delta G_{gas}^{LIG} + \Delta G_{gas}^{REC})] \quad (27)$$

$$\Delta G_{solv} = [\Delta G_{solv}^{COM} - (\Delta G_{solv}^{LIG} + \Delta G_{solv}^{REC})] \quad (28)$$

Where $G_{gas,complex}$ is the free energy of complex formation in gas phase and $G_{gas,ligand}$ and $G_{gas,receptor}$ are the energies necessary to separate ligand and receptor components in vacuum. The in-vacuum binding free energy $\Delta G_{bind,gas}$ is calculated using the molecular mechanics force field used to describe the system during the MD simulation. It is obtained by the sum of internal, electrostatic and van der Waals molecular mechanics interactions:

$$\Delta G_{bind}^{gas} = [\Delta G_{int} - (\Delta G_{vdW} + \Delta G_{el})] \quad (29)$$

The calculation of the ΔG_{solv} is more complicated and corresponds to the free energy change associated with the solvation of the compounds. The $\Delta G_{solv,X}$ is considered for each X component (complex, receptor, or ligand) and it is constituted by a polar and non-polar term.

$$\Delta G_{solv,X} = (\Delta G_{polar,X}^{solv} - \Delta G_{non-polar,X}^{solv}) \quad (30)$$

The polar solvation term is calculated by solving the PB or GB equations which consider the solvent as a high dielectric constant medium and the solute as a set of fixed-point charges embedded in a lower dielectric continuum. This approximation relies on the fact that individual water molecules of the solvent are unimportant in the complex formation reaction, but rather the solvent as a whole provides an environment which affects the behavior of the solute system.

The Poisson-Boltzmann equation derives from the Poisson's equation which describes the electrostatic potential generated by a charge distribution in an environment of dielectric coefficient (relative to the permittivity of free space). The GB implicit solvent model [177] is an approximation of the Poisson-Boltzmann equation as it expresses the polar solvation free energy term of X as a sum over all the atomic charges multiplied for g_{ij}^{GB} :

$$\Delta G_{polar,X}^{solv} = \frac{1}{2} \sum_{i,j \in X} q_i q_j g_{ij}^{GB} \quad (31)$$

The quantity g_{ij}^{GB} is determined by solving the Generalized-Born equation:

$$g_{ij}^{GB} = \left(\frac{1}{\epsilon} - 1 \right) \left[r_{ij}^n + B_{ij} \exp \left(- \frac{r_{ij}^n}{AB_{ij}} \right) \right]^{-\frac{1}{n}} \quad (32)$$

where the parameter B_{ij} depends on the distance from the solute-solvent dielectric boundary of atom i and j and the shape of the entire biomolecule; ϵ is the solvent dielectric constant and r_{ij} is the distance between atom i and j . The constants A and n have preset values. Always under the PB and GB implicit solvent approximations, the non-polar contribution to the solvation energy, $\Delta G_{non-polar}^{solv}$ (Eq. 30) is composed of two terms. The first one is the van der Waals interaction and the second one represents the energy spent to create the solvent cavity occupied by the solute:

$$\Delta G_{non-polar}^{solv} = \Delta G_{non-polar}^{vdW} + \Delta G_{non-polar}^{cavity} = \gamma SASA + \beta \quad (33)$$

The equation shows that the non-polar contribution $\Delta G_{non-polar}^{solv}$ is proportional to the solvent accessible surface area (SASA) of component X , while γ and β are values dependent on the method, PBSA or GBSA, applied [178]. At this point, the program can solve Eq. 26 and give the calculated free binding energy of the system.

The MMPBSA.py package [179] is also capable of computing the entropy via normal mode analysis (NMA) and quasi-harmonic approximation (QH).

Normal Mode Analysis [180] assumes that the potential energy surface (PES) around a minimized structure is purely harmonic. In the AMBER *nmode* implementation, the system is first minimized to a local minimum (where the gradient is zero). The Hessian matrix of mass-weighted second derivatives is then calculated and diagonalized to yield $3N-6$ real vibrational frequencies ν_i . The vibrational entropy is derived by treating each mode as an independent quantum harmonic oscillator. The standard partition function yields the absolute vibrational entropy S_{vib} :

$$S_{vib} = R \sum_{i=1}^{3N-6} \left[\frac{\hbar \nu_i / k_B T}{e^{\hbar \nu_i / k_B T} - 1} - \ln(1 - e^{-\hbar \nu_i / k_B T}) \right] \quad (34)$$

Where R is the gas constant, $\hbar$ is the reduced Planck's constant and k_b is the Boltzmann constant. This method separates internal vibrations from translation and rotation providing an exact entropy for a system confined strictly to a harmonic well.

The Quasi-Harmonic approximation [181] relaxes the strict harmonic requirement by deriving effective frequencies from the fluctuations observed during a molecular dynamics simulation. Instead of the second derivative of the potential energy, AMBER (via cpptraj [182]) diagonalizes the mass-weighted covariance matrix of atomic positions. The eigenvalues λ_m of this matrix are inversely proportional to the square of effective frequencies ω_m allowing the configuration space sampled by the MD to be mapped onto a set of effective harmonic oscillators. This yields the QH configurational entropy:

$$S_{QH} = k_B \sum_{m=1}^{3N-6} \left[\frac{\hbar \omega_m / k_B T}{e^{\hbar \omega_m / k_B T} - 1} - \ln(1 - e^{-\hbar \omega_m / k_B T}) \right] \quad (35)$$

While S_{QH} implicitly captures some anharmonicity by fitting a Gaussian envelope over the sampled energy wells, it constitutes an upper bound on the true entropy, as it is derived under the assumption that the probability distribution of the coordinates follows a multivariate Gaussian form.

To address deviations from this Gaussian assumption, advanced implementations in AMBER analysis can apply higher-order corrections. These include anharmonic corrections ΔS_{ah}^{cl} derived from the non-Gaussian shape of individual mode projections, and pairwise correlation corrections ΔS_{pc}^{cl} that account for coupling between modes:

$$\Delta S_{ah}^{cl} = -k_B \sum_m \int p_m(b_m) \ln \left[\frac{p_m(b_m)}{g_m(b_m)} \right] db_m \quad (36)$$

$$\Delta S_{pc}^{cl} = -k_B \sum_{m < n} \iint p_{m,n}(b_m, b_n) \ln \left[\frac{p_{m,n}(b_m, b_n)}{p_m(b_m)p_n(b_n)} \right] db_m db_n \quad (37)$$

Where p represents the actual probability distribution from the trajectory and g represents the ideal Gaussian reference.

NMA provides an analytically rigorous baseline for stable, rigid structures and allows for the decomposition of motion into distinct, uncorrelated frequencies. It is computationally efficient for small systems once the minimization is complete. However, the harmonic approximation is often too severe for biological macromolecules, as it ignores multiple energy basins and anharmonic motions at room temperature. Furthermore, the Hessian calculation scales as $O(N^3)$ (or $O(N^2)$ with cutoffs), making it computationally prohibitive for very large systems in explicit solvent. It is also extremely sensitive to the quality of the energy minimization; failing to reach a true local minimum results in imaginary frequencies (negative eigenvalues).

On the other hand, QH incorporates solvent effects, temperature dependence, and multiple conformational substates implicitly because it is based on MD sampling. It captures the effective width of the energy wells rather than just the curvature at the bottom, often yielding more realistic entropy values for flexible biomolecules. The principal limitation arises from slow sampling convergence: the covariance matrix exhibits substantially slower convergence than the energy, thereby necessitating significantly extended simulation times to obtain a statistically stable estimate of the entropy. Additionally, if the

molecule visits two distinct basins (e.g., an open and closed state), QH treats this as one giant Gaussian distribution, significantly overestimating the entropy ($S_{\text{QH}} \geq S_{\text{true}}$).

3.2 Grid-based docking and all-atom MD

The initial step of the virtual screening procedure consisted in defining the search boxes used by AutoDock Vina. These boxes delimit the spatial region within which the ligand is allowed to move and are specified in the Vina configuration file. The selection of grid dimensions and positions is necessary for achieving an optimal compromise between computational efficiency and the probability of encompassing all relevant binding sites.

In this study, instead of employing a single box covering the entire graphene oxide sheet, a grid of multiple smaller boxes was defined. An *in-house* script takes as input the GO coordinates and partitions the space according to a user-defined parameter. It then generates n boxes, 9 in this case, for both the top and bottom faces of the GO sheet, resulting in a total of 18 boxes. The box thickness was set to 20 Å to accommodate ligands of varying sizes, ranging from small PFAS to larger antibiotic molecules. This configuration ensures an adequate exploration of the entire GO surface. A visual representation of the grid boxes is provided in [Figure 11](#).

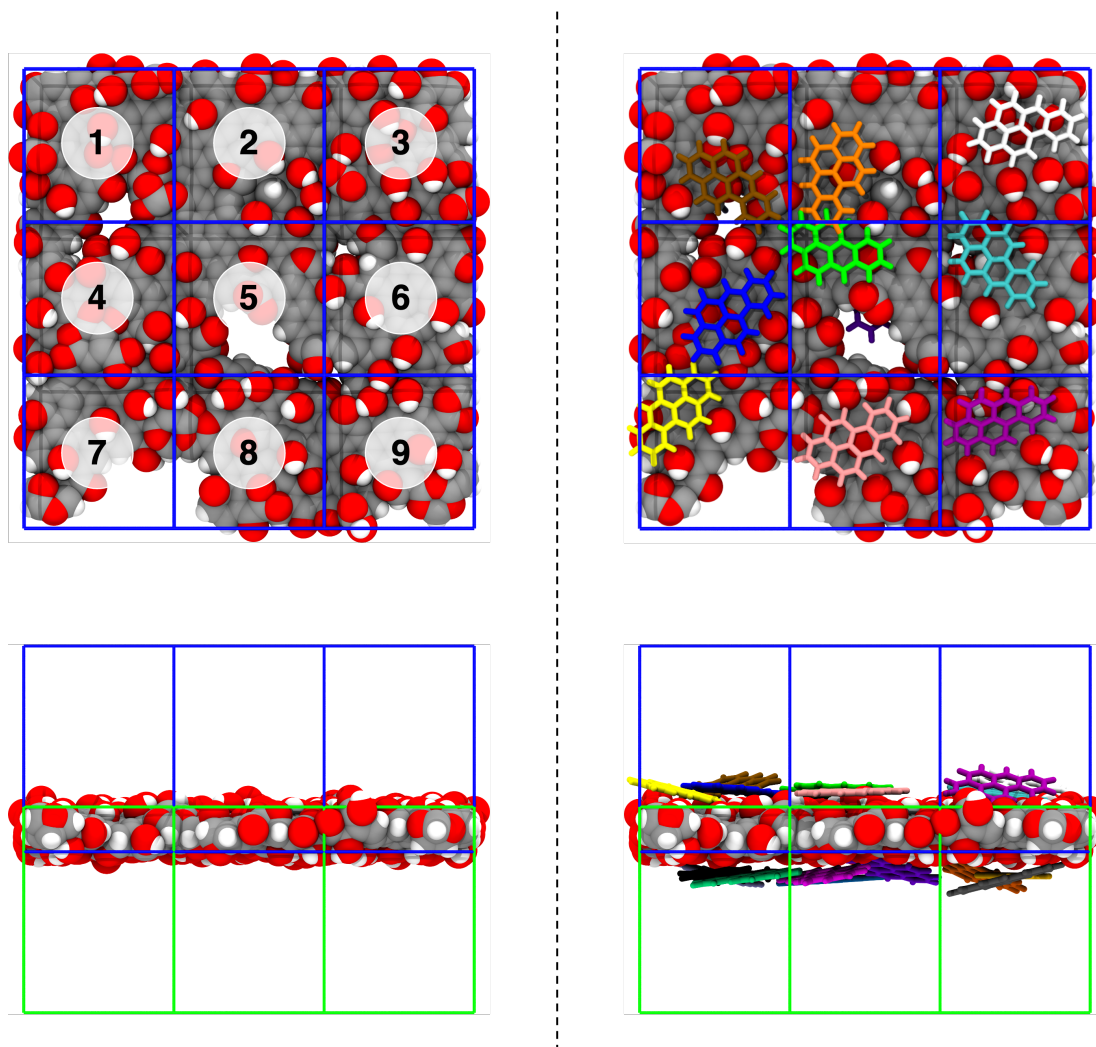


Figure 11: Grid-based docking strategy for graphene oxide surface exploration. (Left) Definition of the nine-section grid of Vina boxes to ensure appropriate sampling. **(Right)** Results following the molecular docking simulations, displaying the optimal binding pose of benzo-a-pyrene within each defined section.

Subsequently, docking parameters were systematically benchmarked to ensure adequate conformational sampling and statistical robustness, given that AutoDock Vina generates binding poses in a stochastic manner. The influence of exhaustiveness, num_modes, and the number of independent runs on docking accuracy and computational cost was evaluated. This analysis identified an efficient parameter set for the systems under study, namely exhaustiveness = 16, num_modes = 10, and number of runs = 3. Increasing these values did not produce substantially different binding poses and only led to a higher computational cost.

Using the optimized parameter set, a total of 540 poses were generated per ligand@GO combination, resulting in approximately 86000 poses for GOX8-P935 and REGOX8-P980 systems combined. While a structural clustering analysis of the generated poses and initial ligands was considered to evaluate the frequency and diversity of binding modes, the extreme scale of the virtual screening campaign and the structural redundancy of the studied drug scaffolds prioritized a high-throughput logic. Therefore, a preliminary filtering step was applied by retaining only the best-scoring pose for each Vina box and each run, such that 2880 poses were ultimately selected for refinement via AA-MD.

Each EC@GO complex was solvated in a cubic box of TIP3P water, ensuring a minimum distance of 20 Å between any solute atom and the box edge. System neutrality was achieved by adding Na⁺ counterions. Energy minimization was performed in two stages: 5000 steps of steepest-descent followed by 5000 steps of conjugate-gradient minimization.

The systems were then equilibrated under NVT and NPT ensembles for 1 ns each prior to production runs, which consisted of 180 ns of simulation in total (10 ns for each of 18 selected poses). As further rationalized in the concluding chapter, this deliberately short 10 ns timescale per pose was necessary to prevent diffusion-induced pose convergence. Extended simulations would have caused ligands to drift across the basal plane and collapse into redundant energy minima, thus defeating the purpose of comprehensively and forcefully sampling the heterogeneous GO surface. Therefore, this timeframe allowed for essential local structural relaxation while preserving the macroscopic spatial distribution generated by the initial grid-based docking. Simulations were conducted under constant temperature and pressure conditions to ensure adequate sampling, with periodic boundary conditions applied throughout all stages.

Equilibration and production simulations were carried out with a 2 fs integration time step. The temperature was maintained at 298 K using a Langevin thermostat with a collision frequency of 1 ps⁻¹, while the pressure was controlled at 1 bar under isotropic conditions. All covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm. The resulting trajectories comprised 1000 frames each and were post-processed with ccpptraj [182], including trajectory wrapping and alignment. Solvent molecules and counterions were removed to reduce data volume, yielding an overall dataset of approximately 2 TB corresponding to ~29 μs of MD simulation time.

3.3 Results

3.3.1 Rationale for selection of binding affinity protocol

The initial investigation was centered on GOX8-P935, which is the most soluble variant within the GO-DB library. The binding affinity of each EC@GOX8-P935 complex was evaluated using the MM-GBSA method in combination with the Onufriev-Bashford-Case Generalized Born Model II (IGB = 5) [177]. This particular implicit solvation model was adopted to maximize the transferability of the computational protocol: IGB = 5 is widely regarded as a reference standard for protein molecular dynamics simulations [183,184], and its use here provides a consistent thermodynamic framework that enables direct comparison between the present environmental screening and future simulations involving graphene oxide-protein interfaces.

Note that for each ligand, the reported binding free energy corresponds to the unweighted mean value obtained from the 18 independent trajectories. While this approach implicitly assumes equal probability for all poses to map the macroscopic interaction propensity across the whole material, it was implemented for two principal practical reasons. First, with respect to surface heterogeneity, sampling the entire GO surface inherently incorporates the material’s topological and chemical diversity. The averaged binding energy thus represents a statistical ensemble of adsorption events, spanning high-affinity hotspots as well as less favorable sites, and therefore yields a more realistic estimate of the global interaction propensity than would be obtained from a single local minimum. Second, regarding stereochemical convergence, the computed binding affinities do not exhibit substantial divergence among different stereoisomers. This observation indicates that the initial docking optimization reliably identifies stable binding modes regardless of chirality, and that each stereoisomer adjusts its conformation to the GO surface with only minor energetic differences.

To further validate this second assumption and to quantitatively assess the influence of stereochemistry on binding affinity, box plots were constructed for each ligand. Figure 12 presents the distribution of binding affinities for ofloxacin and heptachlor across the simulated trajectories. For both compounds, the median binding energies associated with the distinct stereoisomers display no appreciable deviation, indicating that the stereochemical configuration does not induce a pronounced shift in binding

preference under the examined conditions. Moreover, the broad dispersion of the data is primarily attributable to the distinct energetic profiles of the 18 different docking poses on the graphene surface. This finding supports the conclusion that the specific adsorption position and molecular orientation on the sheet exert a greater influence on binding energy than the stereochemical configuration itself (first assumption).

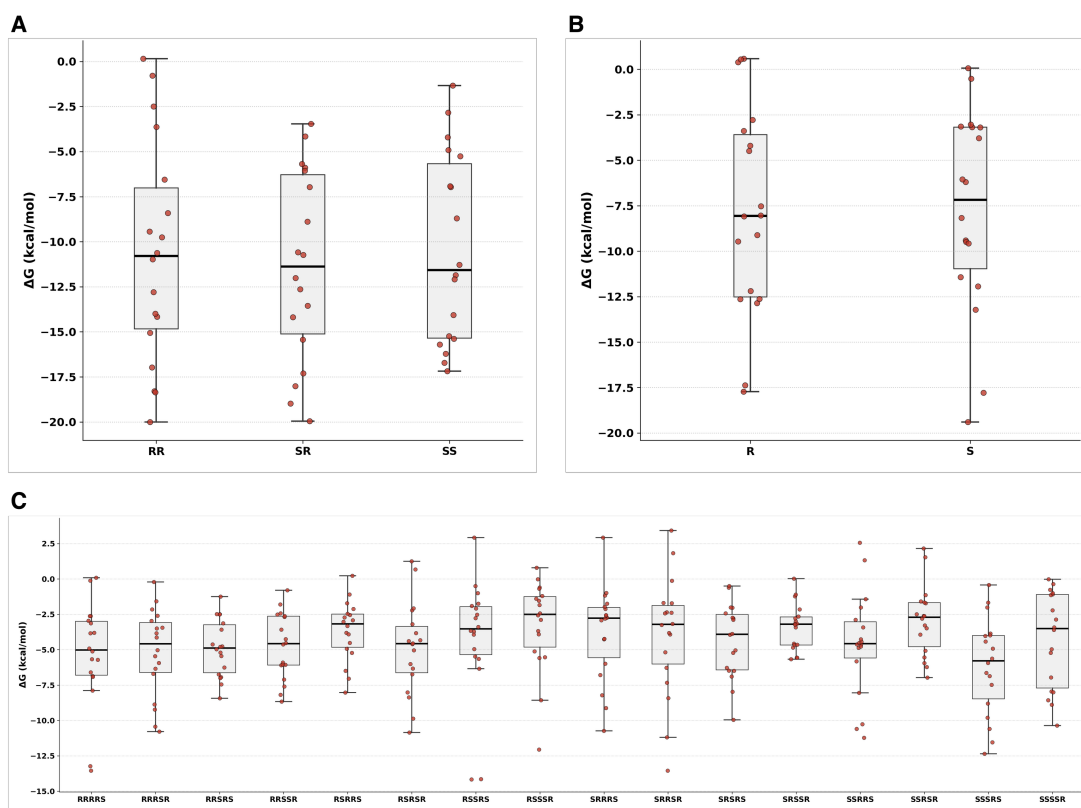


Figure 12: Comparison of binding affinities among stereoisomers. Box plots illustrating binding affinities for the stereoisomers of (A) bis(2-ethylhexyl) phthalate (DEHP), (B) ofloxacin, and (C) heptachlor. Notably, the median binding energy remains relatively stable across the different stereochemical configurations for each compound, indicating comparable interaction profiles regardless of stereochemistry.

Once the comparable behavior of the stereoisomers had been established, each EC@GO complex was ranked according to its binding affinity. However, although MM-GBSA is a widely used and reliable method for ligand ranking [185], it is known to overestimate absolute binding affinities because it does not explicitly account for molecular size, with larger molecules displaying artificially enhanced interaction energies owing to their greater number of atomic contacts with the surface [184,186]. To mitigate this bias and obtain a more realistic description of the adsorption mechanism, the entropy

contribution (ΔS) of the system was evaluated by means of normal mode analysis [180,187].

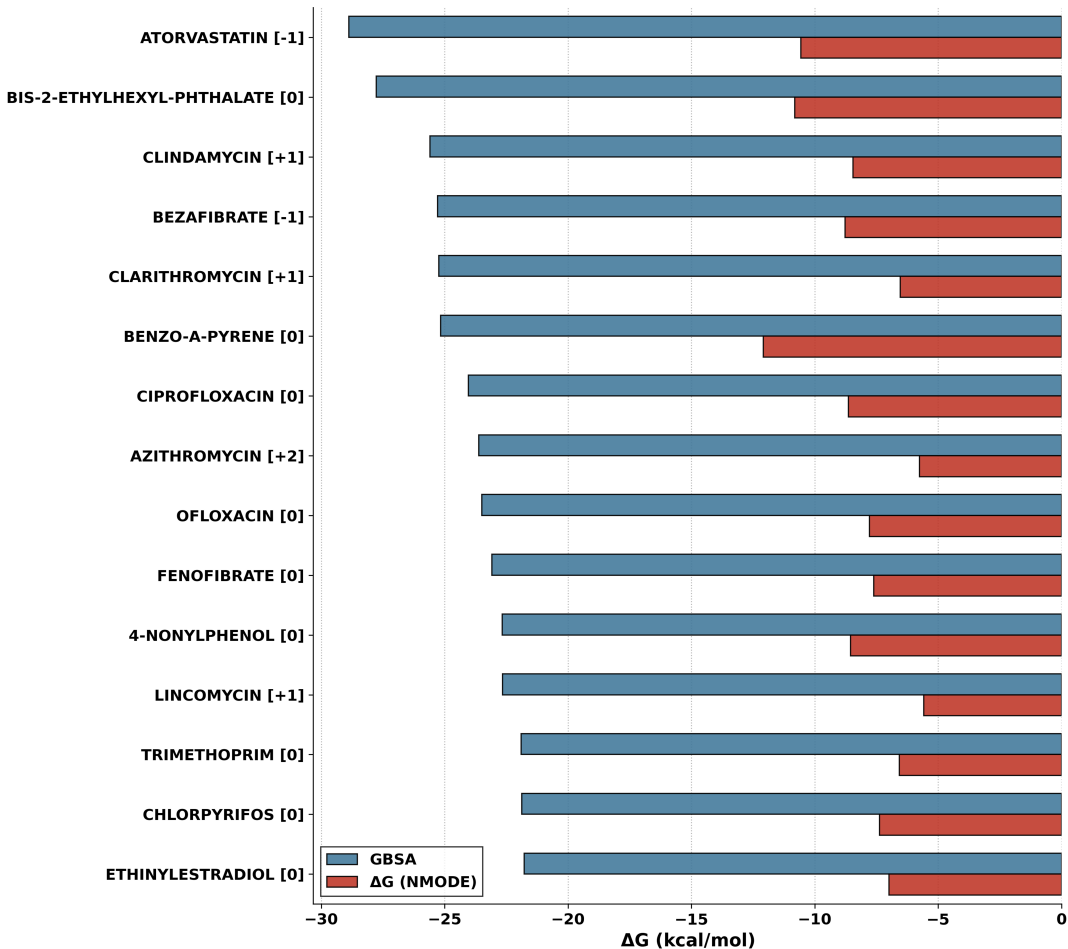


Figure 13: Impact of conformational flexibility and steric bulk on binding affinity. Comparison between standard GBSA and NMODE-corrected binding free energies. Major entropic losses upon binding are primarily driven by pronounced molecular flexibility (e.g., bezafibrate and atorvastatin) or substantial steric bulk (e.g., clarithromycin).

The impact of entropic contributions becomes particularly evident when examining the highest-ranked ligands. Complexes that exhibit the largest entropic penalties are characterized either by pronounced conformational flexibility, as in the cases of bezafibrate and atorvastatin, or by substantial steric bulk, as exemplified by clarithromycin. This behavior is clearly illustrated by the Ligand Efficiency (LE) analysis presented in Figure 14. Defined as $LE = \frac{\Delta G}{N_{HA}}$, where N_{HA} denotes the number of heavy atoms, ligand efficiency provides a size-normalized descriptor of binding affinity [188]. Because a more negative binding free energy (ΔG) reflects a more thermodynamically stable complex, more negative LE values indicate superior

binding efficiency on a per heavy atom basis. This metric is therefore essential for discriminating ligands that achieve high affinity via optimized, specific interactions from those that appear potent primarily due to an increase in molecular size [189–191].

A comparative analysis of the enthalpy-only energy landscape with the entropy-corrected results corroborates the aforementioned conclusion: explicitly accounting for conformational degrees of freedom acts as a filter that disfavors larger and more flexible compounds. As depicted in the scatter plots, where larger marker sizes correspond to higher molecular weight, these bulky ligands generally display inferior ligand efficiency (as indicated by the color scale shifting towards red). Incorporation of entropic contributions thus prevents such molecules from attaining spuriously high rankings solely as a consequence of their size.

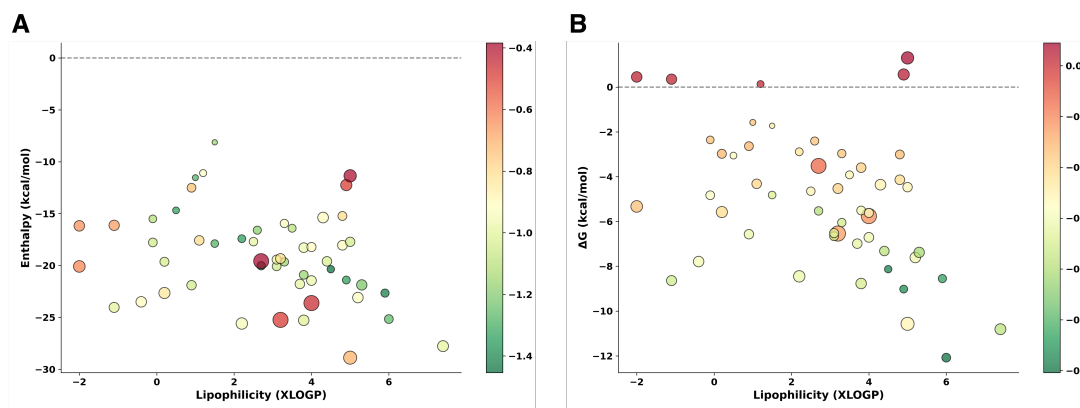


Figure 14: Impact of molecular weight and lipophilicity on ligand efficiency. Scatter plots correlating (A) binding enthalpy and (B) Gibbs free energy with compound lipophilicity (XLOGP). Each data point represents a specific complex, where the marker size is directly proportional to the compound’s molecular weight. The color gradient reflects the Ligand Efficiency (LE), ranging from excellent (green) to poor (red): bulkier ligands (larger dots) predominantly exhibit inferior ligand efficiency.

Up to this point, the reported results have been obtained using the GB-OBC II model. However, given that GO constitutes a non-biological interface with physicochemical properties that differ substantially from those of conventional protein targets, the impact of alternative GB solvation models was systematically investigated. Figure 15 presents a heatmap of the binding free energies for the 53 EC@GOX8-P935 complexes analyzed, colored according to their ΔG values, thereby illustrating how the selection of the solvation model modifies the energetic ranking: the significant variations in binding free energies and rankings highlight a strong model dependency, which is a known

limitation of approximate continuum models when applied to highly charged interfaces. Consequently, absolute energy values must be interpreted with caution, and only macroscopic trends consistent across models should be considered robust.

Inspection of the heatmap supports the exclusion of the GBn models ($\text{IGB} = 7$ and $\text{IGB} = 8$). These models, developed by Mongan, Simmerling, McCammon, Case, and Onufriev [192] and later refined [193], incorporate a pairwise correction term to GB-HCT to approximate a surface dielectric boundary, specifically aimed at eliminating interstitial high-dielectric regions smaller than a solvent molecule. This correction is applied to all atoms and is explicitly dependent on geometry, surpassing the geometry-independent, average rescaling strategy of GB-OBC, which predominantly influences buried atoms. In the present system, GBn produced physically inconsistent results for perfluorinated compounds such as PFOA and PFOS. Experimental evidence indicates that these anionic surfactants experience electrostatic repulsion from the negatively charged GO surface [194,195]; in contrast, $\text{IGB} = 7$ and 8 erroneously rank them among the favorable binders (e.g., $\text{PFOA} \approx -8$ kcal/mol), indicating a failure to correctly estimate either the de-solvation penalty or the electrostatic repulsion component.

With respect to earlier GB models, the Onufriev-Bashford-Case model I ($\text{IGB} = 2$) is rarely employed in contemporary simulation protocols. As the precursor of OBC2 ($\text{IGB} = 5$), it does not include the more refined empirical corrections to effective atomic radii and surface tension that characterize the newer model and generally behaves as a transitional parametrization, offering no clear advantage for the present graphene oxide system.

The Hawkins-Cramer-Truhlar model ($\text{IGB} = 1$) [196], in contrast, induces a pronounced rearrangement of the binding affinity ranking, strongly favoring positively charged compounds (e.g., azithromycin, clarithromycin). Although $\text{IGB} = 1$ is often regarded as less accurate for proteins owing to its simplified treatment of salt screening [177], this behavior may more faithfully reflect the physicochemical reality of the GO interface. As a matter of fact, graphene oxide exhibits a markedly negative zeta potential (approximately -40 mV at neutral pH) [197–199], primarily due to the deprotonation of surface carboxyl groups. Accordingly, the strong electrostatic attraction of cationic pollutants predicted by $\text{IGB} = 1$ is theoretically consistent. If corroborated by subsequent

experimental studies, $\text{IGB} = 1$ could prove particularly suitable for modelling specific GO-cation interactions.

Nevertheless, for the sake of methodological consistency in this study, $\text{IGB} = 5$ (OBC2) was retained as the optimal generalized Born model. This parametrization yields a balanced estimation of solvation energetics, accurately capturing the low binding affinity of anionic molecules while providing a robust and reproducible solvation reference that facilitates direct comparison with future protein-GO simulation studies.

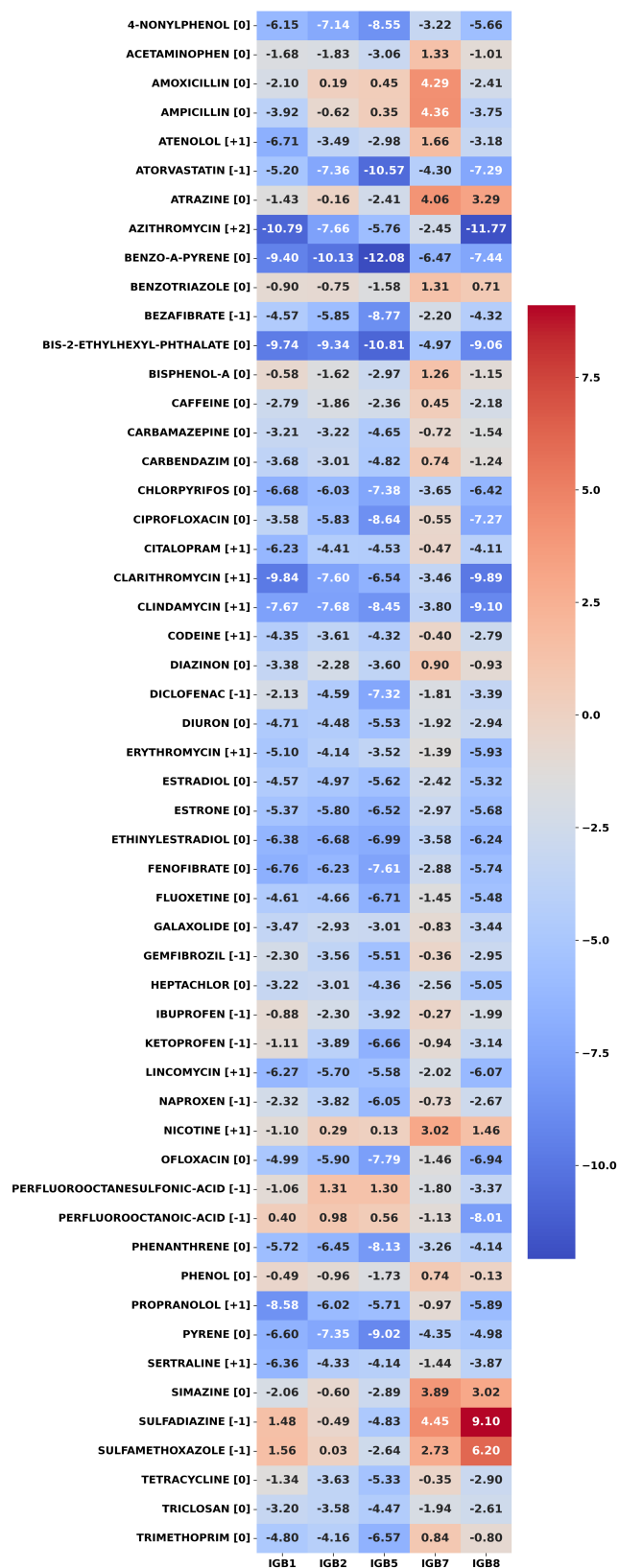


Figure 15: Impact of GB solvation models on binding affinity rankings. The variations in the energetic profiles highlight significant model-dependent behaviors, particularly regarding the handling of highly anionic (e.g., PFOA, PFOS) and cationic (e.g., azithromycin, clarithromycin) compounds at the GO interface.

3.3.2 Final ranking and focus on selected contaminants

The final ranking of all the 53 complexes mediated across their respective stereoisomers is shown in [Figure 16](#).

The computational evaluation of adsorption free energies on GOX8-P935 establishes a distinct hierarchy of affinities, with the 15 highest-ranking candidates exhibiting ΔG values spanning from -12.1 kcal/mol for benzo(a)pyrene to -7.0 kcal/mol for 17 α -ethinylestradiol. This upper subset comprises a chemically heterogeneous set of high-concern contaminants, including planar polycyclic aromatic hydrocarbons such as benzo(a)pyrene, whose leading position is consistent with strong hydrophobic and dispersive interactions that phenomenologically reproduce planar π - π stacking stacking-like geometries, as well as commercially significant pharmaceuticals and industrial plasticizers that are classified as Priority Hazardous Substances under the EU Water Framework Directive (WFD) [\[200\]](#).

Importantly, this thermodynamic ordering is congruent with experimentally established trends reported in the literature; for example, the higher predicted affinity of ciprofloxacin (rank 6, -8.6 kcal/mol) relative to its structural analogue ofloxacin (rank 10, -7.8 kcal/mol) is in agreement with documented sorption behavior [\[201\]](#). Moreover, the model accurately predicts unfavorable or weak interactions for contaminants such as perfluorooctanoic acid (PFOA; 0.6 kcal/mol) and perfluorooctanesulfonic acid (PFOS; 1.3 kcal/mol), as well as for small polar molecules including bisphenol A (-3.0 kcal/mol) and caffeine (-2.4 kcal/mol). These predictions are consistent with literature reports indicating that such species, particularly anionic polyfluoroalkyl substances and highly hydrophilic compounds, frequently require adsorbents based on reduced graphene oxide or amino acid-functionalized GO to sufficiently mitigate electrostatic repulsion and improve hydrophobic interactions [\[65,197,202,203\]](#).

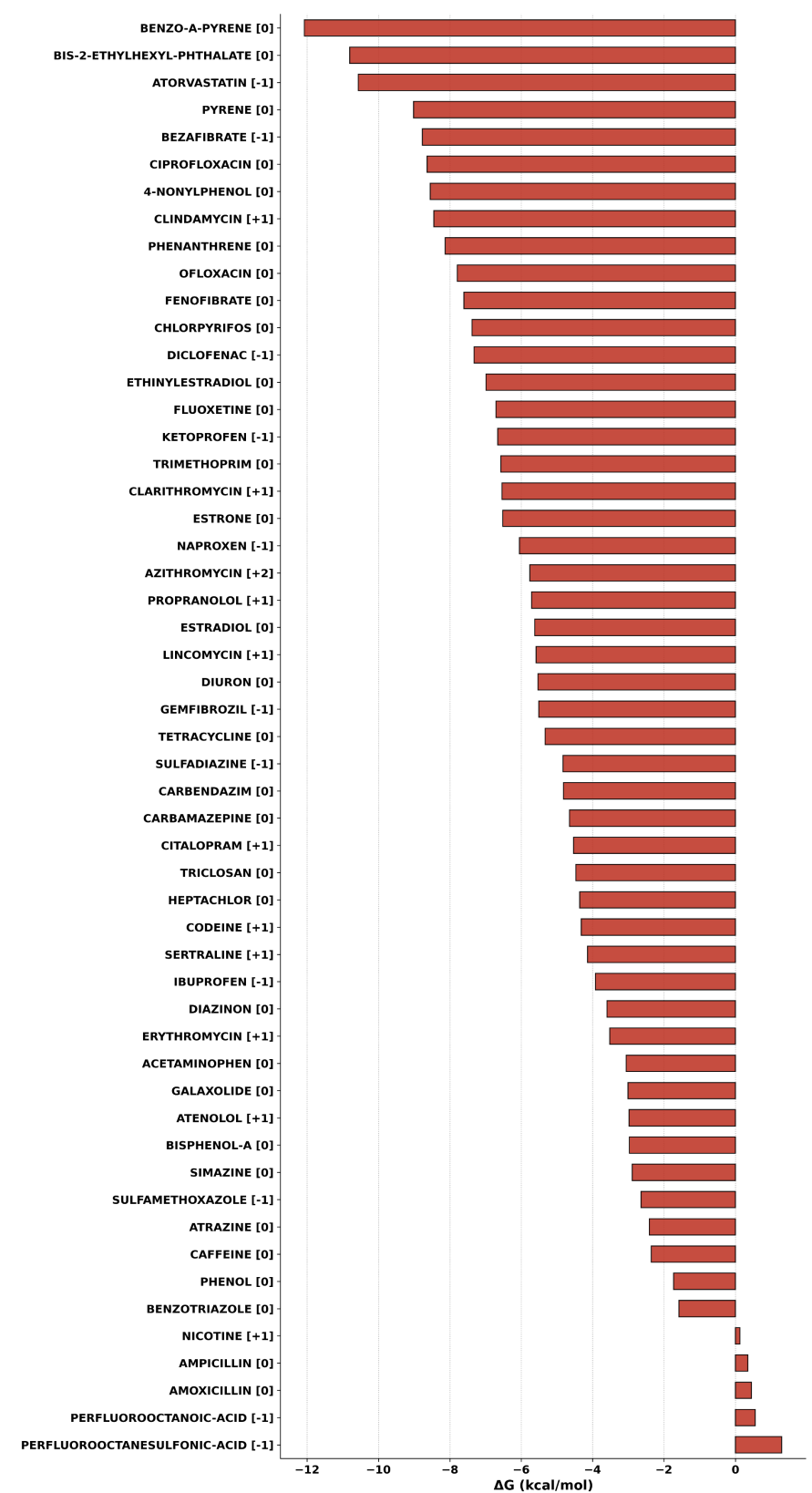


Figure 16: Binding free energy ranking of the screened priority contaminants. The computational ranking establishes a distinct affinity hierarchy, successfully identifying strong binders among priority hazardous substances while correctly predicting the poor adsorption of highly polar or anionic species. The net charge of each pollutant is reported in brackets.

Based on these findings and prioritizing the representation of diverse chemical families rather than a strict adherence to the energetic ranking, bis(2-ethylhexyl) phthalate (DEHP), atorvastatin, and chlorpyrifos were selected as the primary analytes for detailed investigation.

DEHP (rank 2, -10.8 kcal/mol) was chosen as a representative industrial compound and presents a distinct steric challenge due to its commercial occurrence as a complex mixture of stereoisomers. Atorvastatin (Rank 3, -10.6 kcal/mol) was selected as the pharmaceutical representative. Unlike lower molecular weight drugs such as diclofenac (rank 13, -7.3 kcal/mol), this high molecular weight statin constitutes a pseudo-persistent lipophilic contaminant that is marketed as a single, well-defined (3R,5R)-enantiomer, thereby enabling the systematic exploration of stereospecific adsorption phenomena. Finally, the inclusion of chlorpyrifos (rank 12, -7.4 kcal/mol) ensures coverage of organophosphate pesticides: currently under consideration for prioritization in the ongoing WFD revision, this molecule introduces distinct heteroatomic interactions involving chlorine, phosphorus, and sulfur, which are mechanistically different from those governing the adsorption of the primarily aromatic or pharmaceutical compounds described above.

BIS(2-ETHYLHEXYL) PHTHALATE

Bis(2-ethylhexyl) phthalate (DEHP) is a high molecular weight phthalate ester and one of the most extensively used plasticizers in the global polymer industry, predominantly employed to confer flexibility and durability to polyvinyl chloride (PVC)-based materials, including medical devices, consumer products, and construction components. Structurally, DEHP comprises a benzene-1,2-dicarboxylic acid core (phthalic acid) esterified with two branched 2-ethylhexanol moieties. This specific architecture confers pronounced hydrophobicity ($\log P \approx 7.4$) and extremely low aqueous solubility (0.003 mg/L at 25 °C), physicochemical properties that critically determine its environmental fate and transport [\[204\]](#).

A key factor underlying the environmental relevance of DEHP is its mode of incorporation into PVC matrices: it is physically blended rather than covalently bound to the polymer backbone. As a result, DEHP can continuously migrate and leach from finished products into surrounding media via volatilization, abrasion, and release into wastewater, among other processes [\[205\]](#). Because this continuous and widespread emission outpaces its natural degradation rates, DEHP maintains constant environmental background

concentrations and is therefore classified as a pseudo-persistent organic contaminant. Following its emission into aquatic environments, it exhibits a strong propensity to adsorb to suspended particulate organic matter and sediments. This partitioning reduces its bioavailability in the dissolved phase but simultaneously contributes to long term environmental persistence by limiting exposure to conventional degradation pathways [204,206,207].

From a toxicological perspective, DEHP is a well-established endocrine-disrupting chemical (EDC) associated with reproductive and developmental toxicity across multiple species. It primarily exerts anti-androgenic effects by interfering with androgen biosynthesis, notably via disruption of testosterone production [208,209]. Mechanistically, DEHP and its principal bioactive metabolite, mono(2-ethylhexyl) phthalate (MEHP), act as ligands for peroxisome proliferator-activated receptors (PPAR- α and PPAR- γ). Activation of these nuclear receptors is mechanistically linked to peroxisome proliferation and hepatocarcinogenesis in rodent models [210,211], supporting the classification of DEHP by the International Agency for Research on Cancer (IARC) as possibly carcinogenic to humans (Group 2B) [212,213]. In addition to PPAR-mediated pathways, emerging evidence implicates non-PPAR mechanisms, including the induction of oxidative stress, disruption of redox homeostasis, and epigenetic modifications (e.g., DNA methylation and histone modification changes), all of which are associated with adverse outcomes for fertility, embryonic development, and fetal growth [214,215].

In light of its hazardous properties, including endocrine-disrupting potential, reproductive toxicity, and environmental persistence, DEHP is subject to stringent regulatory control. Within the European Union, it is designated as a Priority Hazardous Substance under the Water Framework Directive and is listed as a Substance of Very High Concern (SVHC) under the REACH (Registration, Evaluation, Authorization and Restriction of Chemicals) regulation, reflecting the need for its progressive substitution and strict risk management [216].

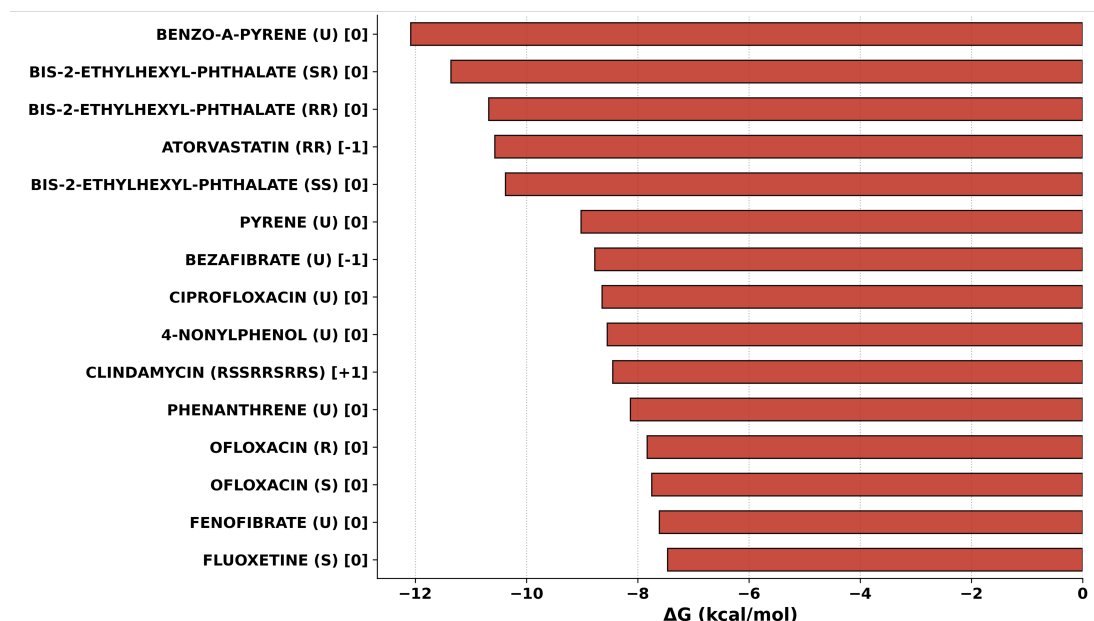


Figure 17: Top 15 binders with individual stereoisomers. All three generated stereoisomers of bis(2-ethylhexyl) phthalate (SR, RR, and SS) consistently position themselves at the top of the hierarchy alongside benzo(a)pyrene, confirming strong and stable binding affinities regardless of their spatial configuration. The net charge of each pollutant is reported in brackets.

Immediately following the polycyclic aromatic compound benzo(a)pyrene, bis(2-ethylhexyl) phthalate displays a distinctively robust adsorption profile, with all three stereoisomers (SR, RR, and SS) occupying high ranking positions (ranks 2, 4, and 5; [Figure 17](#)). This observation highlights the pronounced geometric adaptability of this plasticizer, largely independent of the absolute configuration at its chiral centers. The adsorption process is predominantly governed by non-specific dispersive interactions, reflected in a substantial van der Waals (VdW) contribution of -32.3 kcal/mol. At the structural level, this is achieved through a dual-anchor binding mode, in which the aromatic phthalate core engages in π - π stacking-like interactions with the graphene lattice, while the two bulky, branched 2-ethylhexyl substituents extend over the basal plane to maximize hydrophobic contact area.

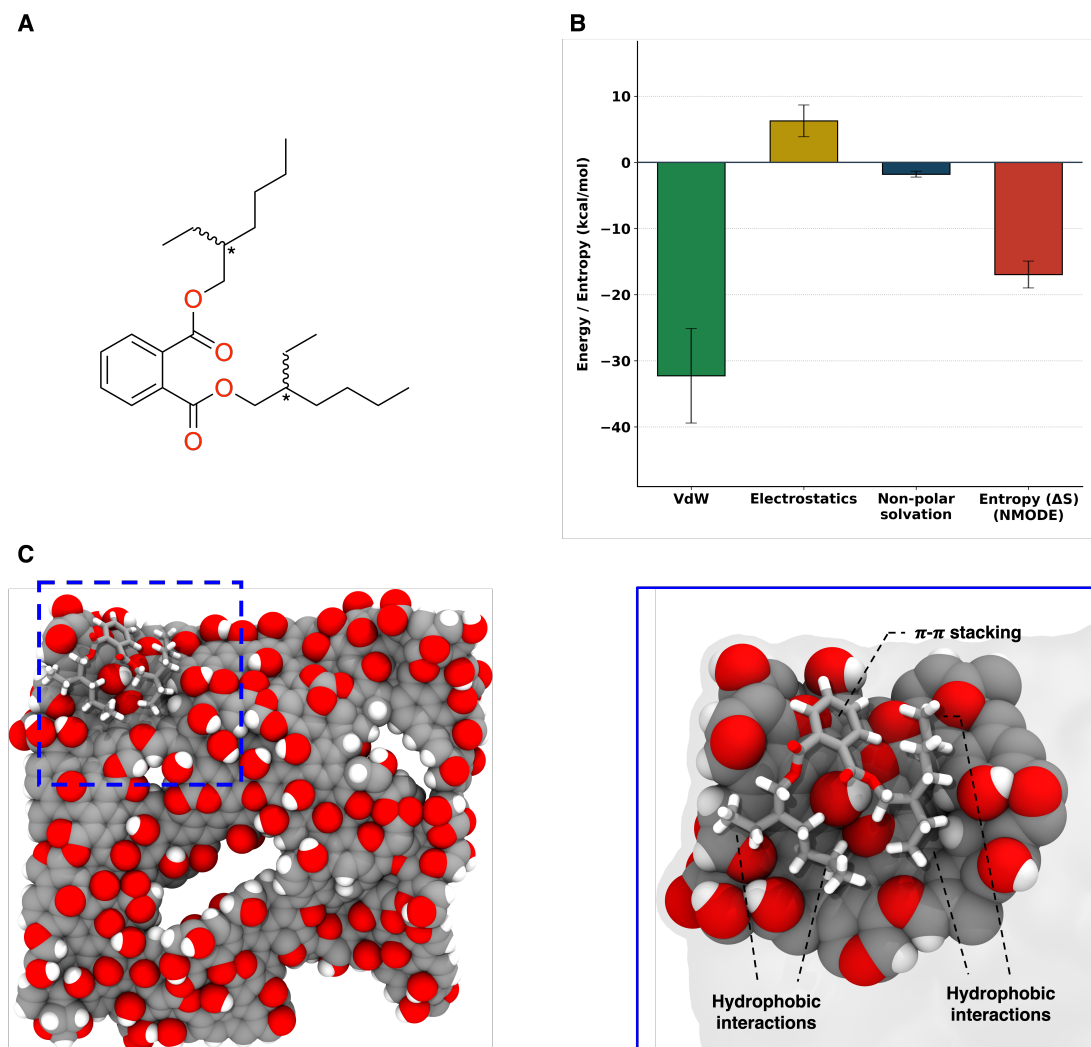


Figure 18: Binding analysis of bis(2-ethylhexyl) phthalate (DEHP). (A) 2D chemical structure of DEHP. (B) Breakdown of the binding free energy components. (C) 3D representation of the best binding pose on the GO surface (left) and a magnified view (right) highlighting key non-covalent contacts. The main binding driving forces are Van der Waals interactions via π - π stacking-like alignment of the central aromatic ring and widespread hydrophobic interactions of the aliphatic chains, which successfully compensate for the significant entropic penalty associated with the molecule's high flexibility. Water and ions were removed for visual clarity.

From an electrostatic standpoint, the net contribution is unfavorable (approximately 6.3 kcal/mol), indicating that the polar ester functionalities do not constitute the primary driving force for adsorption as the energetic penalty associated with de-solvation of these groups exceeds any compensating electrostatic attraction to the surface. Moreover, the similarity in binding energies among the three stereoisomers suggests that the conformational flexibility of the alkyl chains effectively mitigates potential steric constraints imposed by the chiral centers. This adaptive reorganization, however, incurs a pronounced entropic penalty (approximately -17.0 kcal/mol), consistent with

the loss of conformational freedom of the aliphatic tails upon anchoring to the graphene surface.

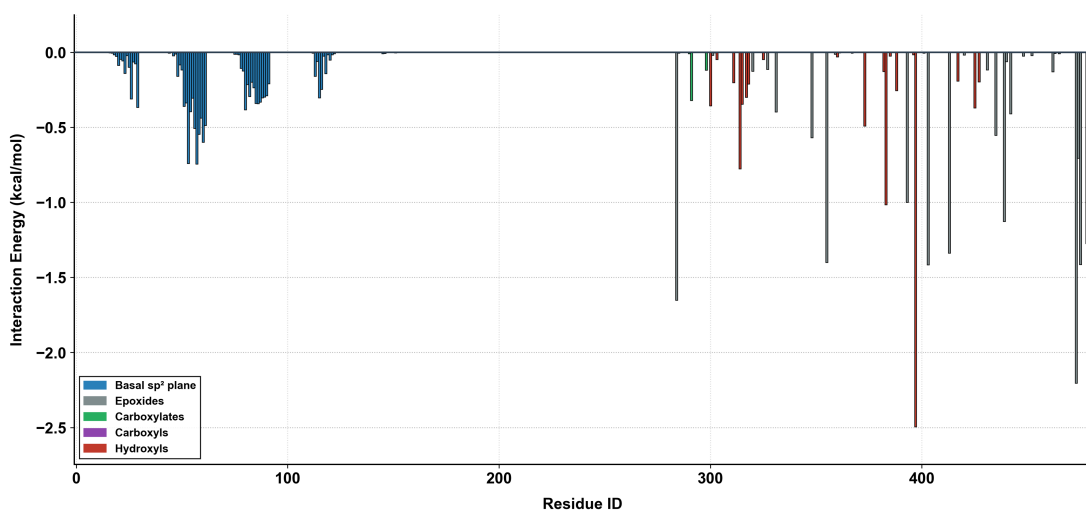


Figure 19: Interaction fingerprint of the DEHP@GO complex. The branched alkyl chains minimize steric strain by interacting with the carbon atoms of epoxide rings and basal domains. Interactions with hydroxyl (red) and carboxylate (green) functionalities further support a geometric model where the bulky ligand prioritizes the material’s perimeter or defect sites.

To obtain a more granular view of these contributions, the interaction fingerprint analysis provides a detailed decomposition of the contact interface arising from the fragment-based parametrization of the GO model (Figure 19). The resulting spectrum exhibits a multimodal distribution, with contributions from the basal sp^2 plane (blue bars) and epoxide groups (grey bars) indicating extensive surface coverage by the branched alkyl chains. Closer examination of the epoxide signals supports the proposed geometric binding model: the lipophilic tails minimize direct contact with the protruding oxygen atoms and instead interact preferentially with the carbon atoms of the epoxide rings, thereby exploiting local lattice distortions (buckling) to maximize dispersive adhesion.

Analogously to the epoxide groups, the prominent peaks associated with surface hydroxyl functionalities (red bars) arise predominantly from interactions with partially positively polarized carbon atoms within hydroxyl-rich regions and do not correspond to the formation of directional hydrogen bonds.

Furthermore, the interaction with edge carboxylate groups (green bars) demonstrates that the molecule preferentially occupies sheet boundaries or

defect sites. However, the lipophilic phthalate interacts mainly with the basal carbon framework associated with these edge groups, rather than with the anionic COO^- termini. This edge selectivity enables the bulky molecule to reduce steric strain on the flat basal plane while taking advantage of the structural defects and accessible carbon scaffolds present at the material's perimeter.

ATORVASTATIN

Atorvastatin is a fully synthetic inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase belonging to the statin class and is currently among the most frequently prescribed pharmaceuticals worldwide for the treatment of hypercholesterolemia and the reduction of cardiovascular risk [217]. From a chemical standpoint, it is a high molecular weight (558.6 g/mol) heptanoic acid derivative featuring a characteristic fluorophenyl substituent and multiple fused and non-fused aromatic rings. In contrast to smaller, more polar pharmaceuticals such as diclofenac, atorvastatin exhibits pronounced lipophilicity ($\log P \approx 6.3$) and very low aqueous solubility. At environmentally and physiologically relevant pH values, atorvastatin is present predominantly in its anionic form ($\text{pK}_a \approx 4.5$). Unlike many industrial contaminants that are marketed as racemic mixtures, atorvastatin is manufactured and used exclusively as the single (3R,5R)-enantiomer [218].

Due to its extensive global consumption and incomplete biotransformation in humans and other organisms, atorvastatin behaves as a pseudo-persistent organic contaminant in aquatic environments. It is routinely detected in wastewater treatment plant effluents and in receiving surface waters at concentrations sufficient to elicit chronic toxic effects in aquatic biota [219,220]. These effects are primarily associated with the disruption of lipid biosynthesis pathways in non-target organisms and encompass a range of severe physiological impairments, including oxidative stress, reproductive dysfunction, and histological evidence of muscle damage (myotoxicity) in both fish and invertebrate species [221,222]. As a result of these concerns, atorvastatin has been included on the U.S. Environmental Protection Agency's Contaminant Candidate List (CCL) and is subject to increasing regulatory and scientific scrutiny within the European Union, thereby constituting a high priority target for the development and implementation of advanced tertiary wastewater treatment technologies [12,223].

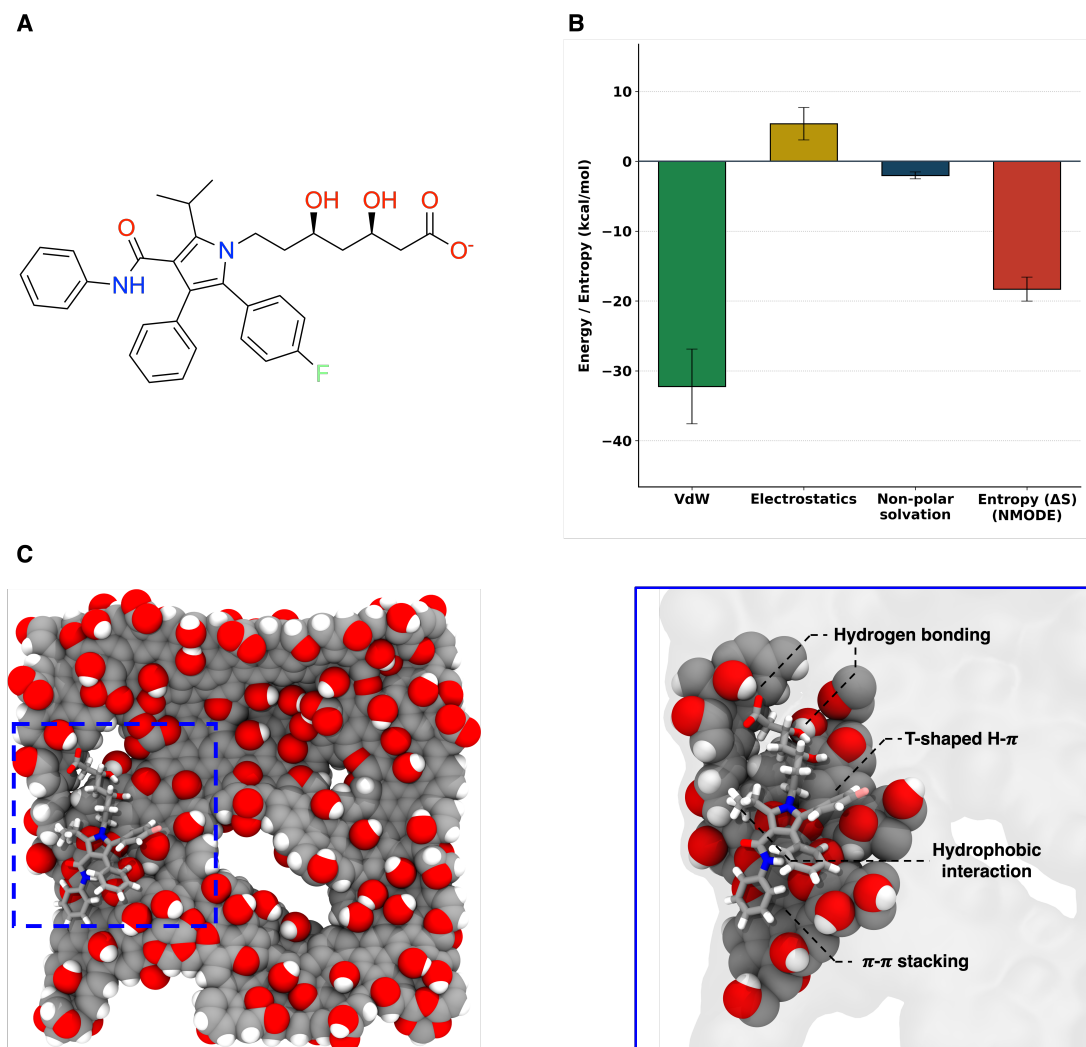


Figure 20: Binding analysis of Atorvastatin. (A) 2D chemical structure of atorvastatin. (B) Breakdown of the binding free energy components. (C) 3D representation of the best binding pose on the GO surface (left) and a magnified view (right) highlighting key non-covalent contacts. The main binding driving forces are Van der Waals interactions via π - π stacking, hydrophobic interactions and a well-defined hydrogen bonding network. Water and ions were removed for visual clarity.

Detailed structural analysis of the atorvastatin@GO complex indicates that the binding mechanism is predominantly governed by hydrophobic and van der Waals interactions, as quantitatively evidenced by the dominant van der Waals contribution of -32.2 kcal/mol. At the molecular level, this high binding affinity is primarily attributed to the phenylcarbamoyl moiety, which engages in a stable π - π stacking-like interaction with the unoxidized sp^2 domains of the GO lattice. In parallel, the bulky isopropyl substituent enhances dispersive interactions by increasing the effective contact area, in conjunction with the aliphatic heptanoic acid chain.

This strong surface adhesion is associated with a considerable thermodynamic penalty in terms of conformational entropy, with the complex exhibiting an entropic contribution of approximately -18.3 kcal/mol. The entropic loss is attributable to the restriction of conformational flexibility of atorvastatin upon adsorption: the highly flexible dihydroxyheptanoic acid side chain and the rotatable isopropyl group, which display mobility in bulk solution, become constrained into a relatively rigid, surface-bound conformation to optimize intermolecular contacts.

Despite this entropic cost, the pharmaceutical establishes a well-defined hydrogen bonding network, driven by the specific stereochemical arrangement of the drug's hydroxyl groups and the terminal carboxylate anion, which act as hydrophilic anchors forming directional bonds with the oxygenated GO edges. However, the localized stabilizing effect of these interactions is ultimately overwhelmed by the global electrostatic repulsion between the atorvastatin anion and the highly negatively charged surface of the graphene oxide. As a result, the net electrostatic contribution is destabilizing, amounting to approximately 5.3 kcal/mol.

Consistent with these observations, interaction fingerprint analysis ([Figure 21](#)) reveals a clear functional specialization among the surface groups. The resulting interaction profile displays high-intensity peaks associated with surface hydroxyls, pinpointing the primary sites responsible for the strong, directional hydrogen bonds described above. In contrast, the broad distribution of favorable contacts with epoxide groups reflects a more subtle geometric mode of interaction: instead of engaging directly with the epoxide oxygen atoms, the ligand interacts predominantly with the carbon atoms of the epoxide ring. Because the epoxide oxygens often protrude toward the opposite face of the basal plane or are sterically shielded, atorvastatin preferentially aligns along the exposed, electron-deficient carbons of these strained rings. In doing so, it exploits the local sp^3 -induced buckling of the GO lattice to maximize dispersive adhesion.

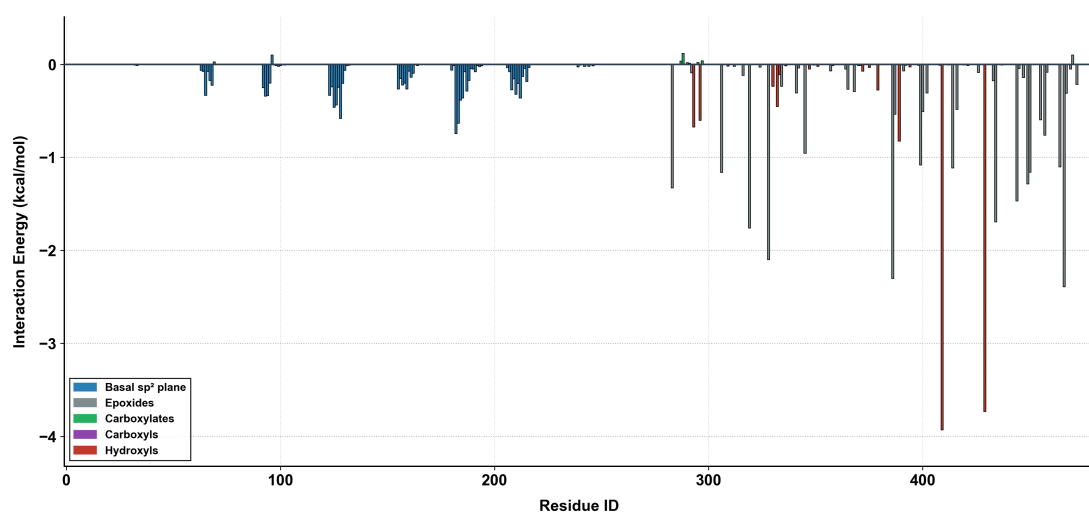


Figure 21: Interaction fingerprint of the atorvastatin@GO complex. High-intensity peaks associated with surface hydroxyls (red) dictate specific hydrogen bond formation, and broader epoxide signals (grey) reflect dispersive interactions driven by the ligand's alignment with sp^3 -hybridized buckled regions of the carbon lattice.

Finally, with respect to the specific contribution of the halogen substituent, the fluorophenyl moiety does not engage in direct coordination or adsorption onto the surface but instead remains oriented toward the bulk solvent phase. Simultaneously, the remaining aromatic rings adopt a staggered conformation to alleviate steric congestion. This arrangement delineates a binding mode governed by an equilibrium between the maximization of π -mediated surface interactions and the minimization of steric repulsion.

CHLORPYRIFOS

Chlorpyrifos is a broad-spectrum organophosphate insecticide that, until recently, represented a key component of global agricultural pest control strategies. From a structural perspective, it is an achiral thiophosphoric acid ester bearing a trichlorinated pyridinyl moiety. This specific pattern of halogenation, together with the central phosphorus-sulfur functional group, confers moderate lipophilicity ($\log P \approx 5$) and low aqueous solubility (~ 1.4 mg/L) [224]. These physicochemical properties promote preferential sorption to organic sediments and suspended particulate matter rather than partitioning into the dissolved aqueous phase [225,226].

From a toxicological standpoint, chlorpyrifos is a potent neurotoxicant that exerts its primary mode of action through irreversible inhibition of acetylcholinesterase (AChE). This effect is frequently enhanced *in vivo* by oxidative desulfuration to its more toxic oxon metabolite, chlorpyrifos-oxon,

which exhibits a higher affinity for AChE and results in sustained cholinergic overstimulation in non-target organisms. Growing toxicological and epidemiological evidence has associated chlorpyrifos exposure, even at relatively low doses, with genotoxic effects and developmental neurotoxicity in humans [225,227–230].

In response to these concerns, the regulatory status of chlorpyrifos has undergone substantial revision. In 2020, the European Commission adopted Implementing Regulation (EU) 2020/17, which non-renewed the approval of chlorpyrifos, effectively prohibiting its use within the European Union [231]. Nevertheless, its environmental persistence, continued detection in agricultural runoff, and accumulation in environmental compartments sustain its relevance as a contaminant of emerging concern. Accordingly, chlorpyrifos has been identified as a candidate substance for inclusion in the Priority Substances list under the ongoing revision of the EU Water Framework Directive, a status that implies the need for advanced treatment and remediation technologies to achieve the proposed ultra-low Environmental Quality Standards (EQS) [232].

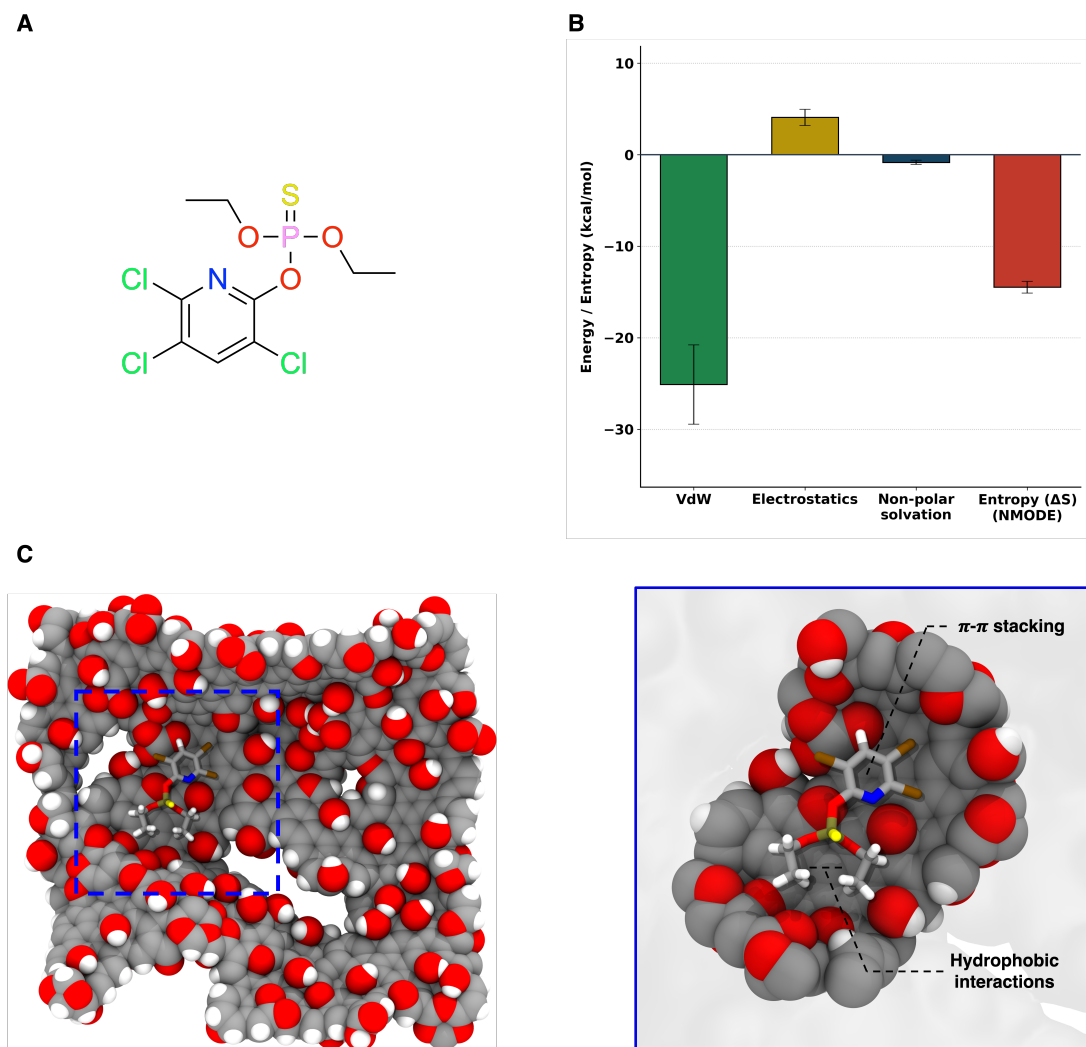


Figure 22: Binding analysis of chlorpyrifos. (A) 2D chemical structure of chlorpyrifos. (B) Breakdown of the binding free energy components. (C) 3D representation of the best binding pose on the GO surface (left) and a magnified view (right) highlighting key non-covalent contacts. The parallel orientation of the 3,5,6-trichloro-2-pyridinyl ring facilitates π - π stacking alignments, alongside secondary hydrophobic stabilization provided by the ethyl substituents. Water and ions were removed for visual clarity.

Consistent with previous findings, the adsorption profile of chlorpyrifos is governed predominantly by non-specific dispersive interactions, as indicated by a van der Waals contribution of -25.1 kcal/mol. The geometric core of this interaction is the 3,5,6-trichloro-2-pyridinyl ring, which adopts an orientation parallel to the basal plane, thereby promoting pronounced π - π stacking, while the ethyl substituents of the phosphate group furnish additional, secondary stabilization through non-specific contacts with the GO surface. Consequently, the three chlorine atoms substituting the aromatic ring are positioned parallel to the graphene plane, which effectively prevents their involvement in specific, directional bonding interactions.

From an electrostatic standpoint, the overall contribution is slightly unfavorable, with a net value of approximately 4.1 kcal/mol, confirming that adsorption is not driven by electrostatic forces as the modest favorable electrostatic attraction is offset by the energetic penalty associated with desolvation of the polar phosphorothioate moiety. Finally, NMODE analysis yields an entropic penalty of -14.5 kcal/mol, indicative of a loss of conformational freedom, particularly of the rotatable phosphoester bonds, required to immobilize the conformationally flexible pesticide on the GO surface.

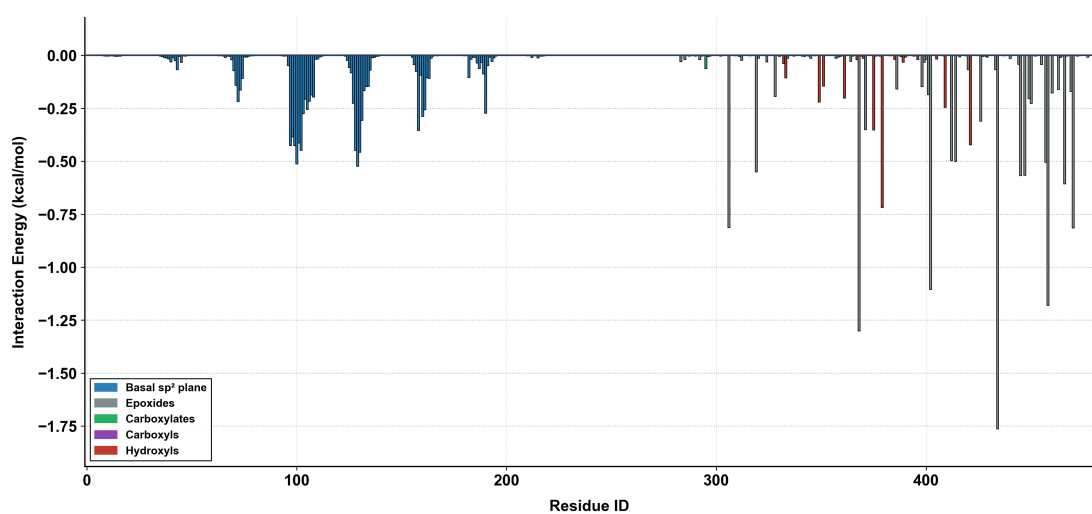


Figure 23: Interaction fingerprint of the chlorpyrifos@GO complex. The multimodal fingerprint highlights the strong preference of the pesticide for π - π stacking of the aromatic ring on non-oxidized graphitic domains, with epoxides and hydroxyls contributing only marginally.

Similarly to the interaction profile of DEHP, the chlorpyrifos fingerprint (Figure 23) is predominantly characterized by contacts with the basal sp^2 -hybridized carbon plane. The high density and amplitude of these peaks indicate that the aromatic trichloropyridine ring preferentially resides above the non-oxidized graphitic domains of the lattice, thereby maximizing π - π stacking and dispersive (van der Waals) interactions. Although favorable contacts with epoxide functionalities and hydroxyl groups are also detected, their frequency is comparatively lower than those associated with the basal carbon signal. This distinctive interaction pattern supports the proposed binding model: chlorpyrifos adsorption is primarily governed by the intercalation and planar alignment of its aromatic core with the extended carbon framework, whereas oxygenated functional groups contribute only secondarily and incidentally to the overall adsorption mechanism.

3.3.3 Comparison with REGOX8-P980

The same protocol was applied to the reduced graphene oxide (rGO) REGOX8-P980, selected as it is most frequently employed in subsequent surface functionalization studies. Analysis of the top 15 complexes indicates an overall more favorable binding affinity for the reduced surface (Figure 24):

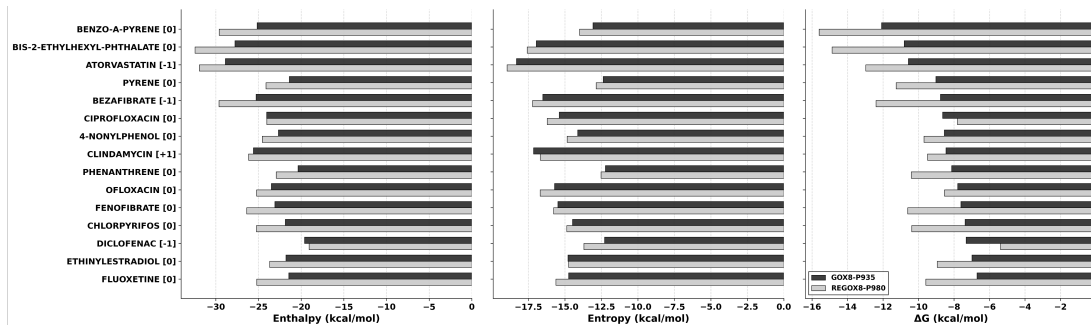


Figure 24: Influence of graphene oxide reduction on binding thermodynamics. While the entropic cost remains relatively stable regardless of the surface (middle panel), REGOX8-P980 provides significantly larger enthalpic stabilization (left panel) due to the restored graphitic domains which promote stronger hydrophobic interactions.

Decomposing the free energy into its thermodynamic components reveals that this enhanced affinity is almost exclusively enthalpy driven. As expected, the entropic penalty remains largely comparable between the oxidized (GOX8-P935) and reduced (REGOX8-P980) systems, reflecting similar losses in the GOs degrees of freedom upon binding. In contrast, the enthalpic term exhibits a marked increase in magnitude (more negative values) for the REGOX system. This enthalpic gain can be mainly attributed to the restoration of the basal sp^2 carbon, which provides a larger, uninterrupted and regular hydrophobic surface area, thereby maximizing van der Waals forces and stacking interactions. Consequently, while the general binding tendency improves, the final thermodynamic ranking of the molecule's shifts depending on the surface composition and geometry. Some molecules, such as bis(2-ethylhexyl) phthalate, pyrene and benzo(a)pyrene, maintain a high affinity for both systems. Given their neutral nature, the adsorption mechanism for these compounds is heavily dominated by hydrophobic interactions, in which the aromatic ring systems, alongside bulky non-polar aliphatic chains, maximize surface contact with the restored sp^2 carbon domains [233].

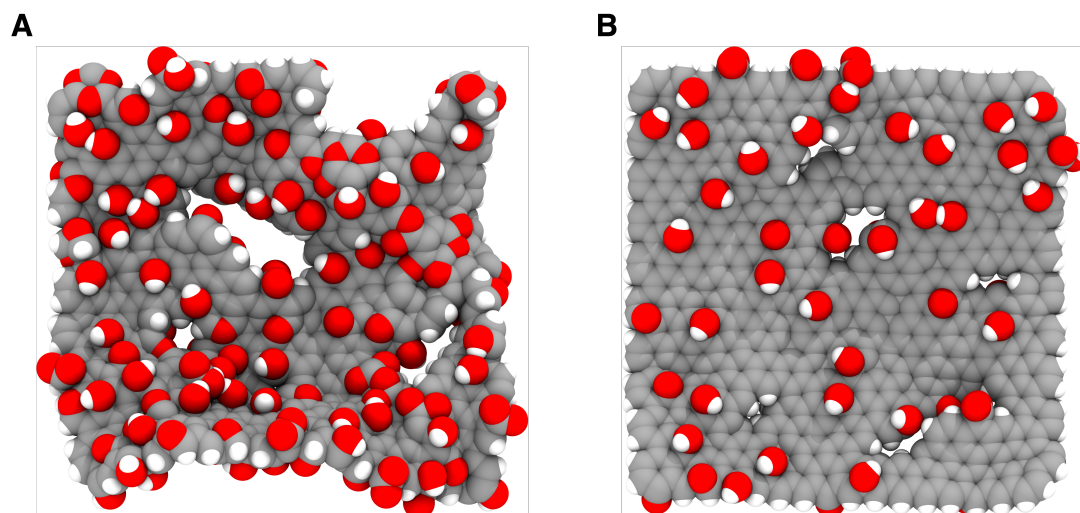


Figure 25: Structural impact of surface functionalization on graphene oxide topology. (A) The extensive presence of epoxides and hydroxyls on the GOX8 surface causes severe lattice distortion, creating a corrugated and jagged interface. (B) The partial removal of these functional groups in the REGOX8 system restores a more continuous, planar carbon framework, facilitating extended π - π stacking interactions.

However, the transition to the reduced surface disproportionately favors specific positively charged compounds, namely azithromycin and propranolol. This improvement is primarily driven by geometric factors. For propranolol, similar to other predominantly aromatic molecules, its aromatic ring can align optimally and lay flat against the restored, planar sp^2 domains of the REGOX surface. In contrast, the highly functionalized GOX surface is much more corrugated and jagged, structurally disrupting this ideal planar contact [199,234]. For azithromycin, the geometric accommodation is twofold: not only does the bulky drug position itself more favorably on the flatter reduced surface, but it also identifies an optimal binding pocket where it establishes multiple hydrogen bonds that are sterically hindered or partially absent on the GOX model.

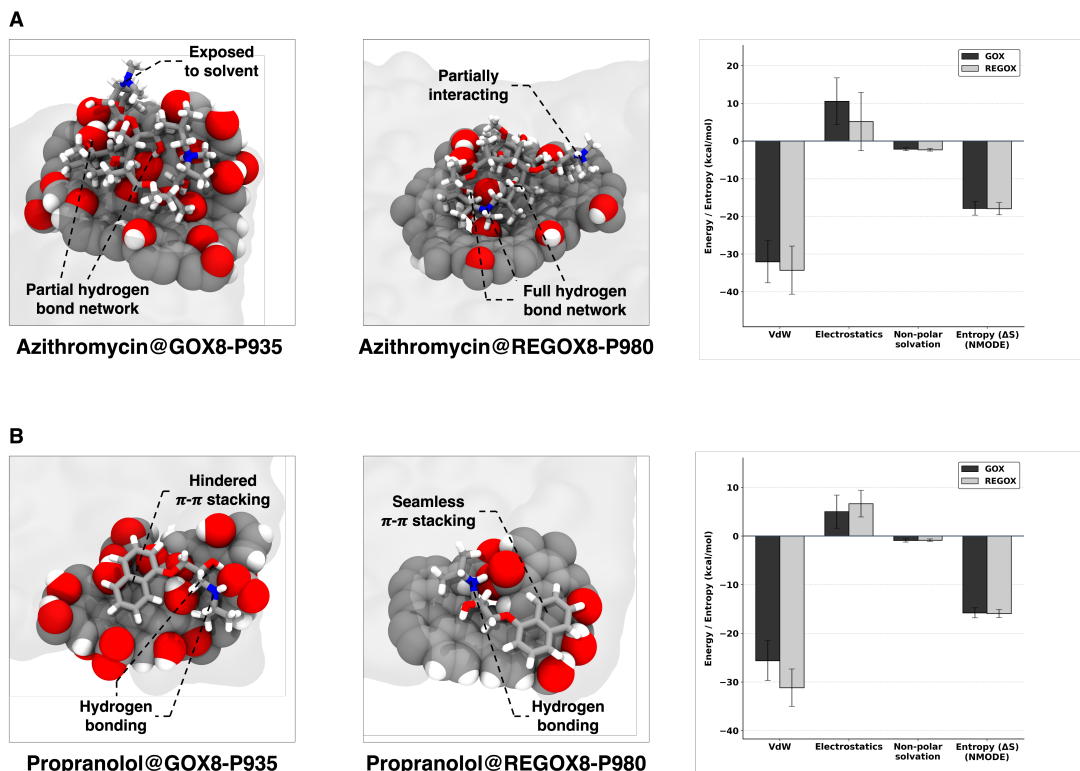


Figure 26: Impact of surface topology on the binding geometry and energetics of azithromycin and propranolol. Atomistic representations and thermodynamic breakdowns for (A) azithromycin and (B) propranolol adsorbed onto GOX (left) and REGOX (middle) models. Structurally, the planar domains of REGOX allow propranolol to lay flat and enable azithromycin to secure a sterically unhindered binding pocket for optimal hydrogen bonding, effectively overcoming the limitations imposed by the corrugated GOX interface.

Despite these geometric rationalizations, the massive enthalpic gain observed for azithromycin remains somewhat counterintuitive. Given its positive net charge, one would expect a stronger electrostatic attraction, and consequently a higher binding affinity, toward the heavily oxygenated GOX surface, which experimentally possesses a much more negative zeta potential [197–199].

This discrepancy between the expected electrostatic behavior and the calculated interaction energy likely stems from the inherent limitations of the *in silico* model itself, specifically its reliance on the random Lerb-Klinowski model [53]. Because this approach distributes functional groups randomly across the basal plane and edges, a single simulated sheet represents only one of countless possible surface microstates. Therefore, a prospective solution for future studies would involve averaging the thermodynamic results not only across multiple binding poses on a single nanosheet, but also across an ensemble of different models generated for the exact same oxidation state. Varying the spatial distribution of defects and functional groups across multiple replicas

would yield a statistically robust average, far better representing the experimental macroscopic reality where thousands of inherently heterogeneous graphene sheets interact simultaneously.

3.3.4 SAR analysis and topological profiling

To transition from the evaluation of individual molecular insights to the elucidation of systemic adsorption rules, the computed binding free energies were coupled with a structural clustering analysis. The primary objective was to ascertain whether the thermodynamic behaviors observed for specific contaminants (e.g., DEHP, atorvastatin) were representative of broader chemical classes, or merely isolated cases.

To achieve this, the entire molecular library was subjected to a topological profiling based on the computation of Morgan Fingerprints (radius 2, 1024 bits). The resulting high-dimensional structural data was subsequently condensed using Multidimensional Scaling (MDS) [235,236] to visualize the chemical space, and partitioned into homogeneous families using the DBSCAN [237] density-based clustering algorithm. A stringent similarity threshold based on the Tanimoto distance [238,239] (> 0.4 , 40% similarity) was imposed. This restrictive criterion was selected to prevent the artificial grouping of structurally dissimilar molecules, thereby ensuring the chemical purity and the thermodynamic consistency of the identified clusters.

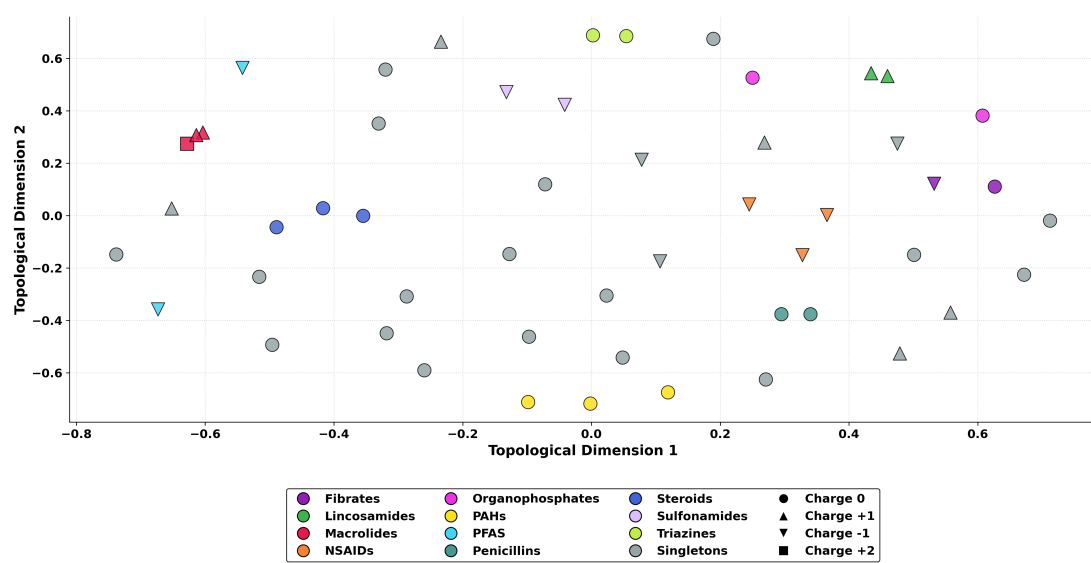


Figure 27: 2D visualization of the structural diversity within the pollutant dataset. High-dimensional topological data (Morgan fingerprints) was condensed using Multidimensional Scaling (MDS) and grouped via the DBSCAN algorithm based on Tanimoto similarity. Distinct chemical classes are differentiated by color, whereas structurally unique compounds that did not cluster are classified as grey singletons. Marker shapes indicate the net formal charge of each contaminant.

As an inevitable consequence of this structural filtering, approximately 50% of the dataset did not cluster and was classified as singletons (in gray, Figure 27), representing structurally unique scaffolds rather than mere topological noise. This outcome does not indicate a methodological flaw of the DBSCAN clustering procedure but instead mirrors the chemical heterogeneity of the environmental library constructed for this study. Because the priority pollutants dataset comprises 53 substances spanning highly diverse therapeutic and industrial classes, specific molecular architectures – such as statins (atorvastatin), complex plasticizers (DEHP), or polycyclic musks (galaxolide) – naturally lack structural analogs within this focused sample.

Notably, the isolation of these compounds into singletons perfectly corroborates the molecular analyses presented earlier. For instance, the dual-anchor binding mode of DEHP – relying on highly flexible, branched alkyl chains – and the complex, sterically demanding structure of atorvastatin are topologically unique within this library. Yet, precisely because of these highly specialized interaction patterns, these singletons feature some of the strongest binding affinities in the entire dataset (e.g., DEHP at -10.8 kcal/mol and atorvastatin at -10.6 kcal/mol). By isolating these unique compounds, the algorithm guarantees that the averaged thermodynamic properties reported for the multi-molecule clusters represent genuine Structure-Activity Relationship (SAR)

trends, devoid of statistical artifacts induced by forced classifications. The integration of the clustering output with the MM-GBSA binding free energies allows for a comprehensive mapping of the adsorption hierarchy across distinct pharmacological classes.

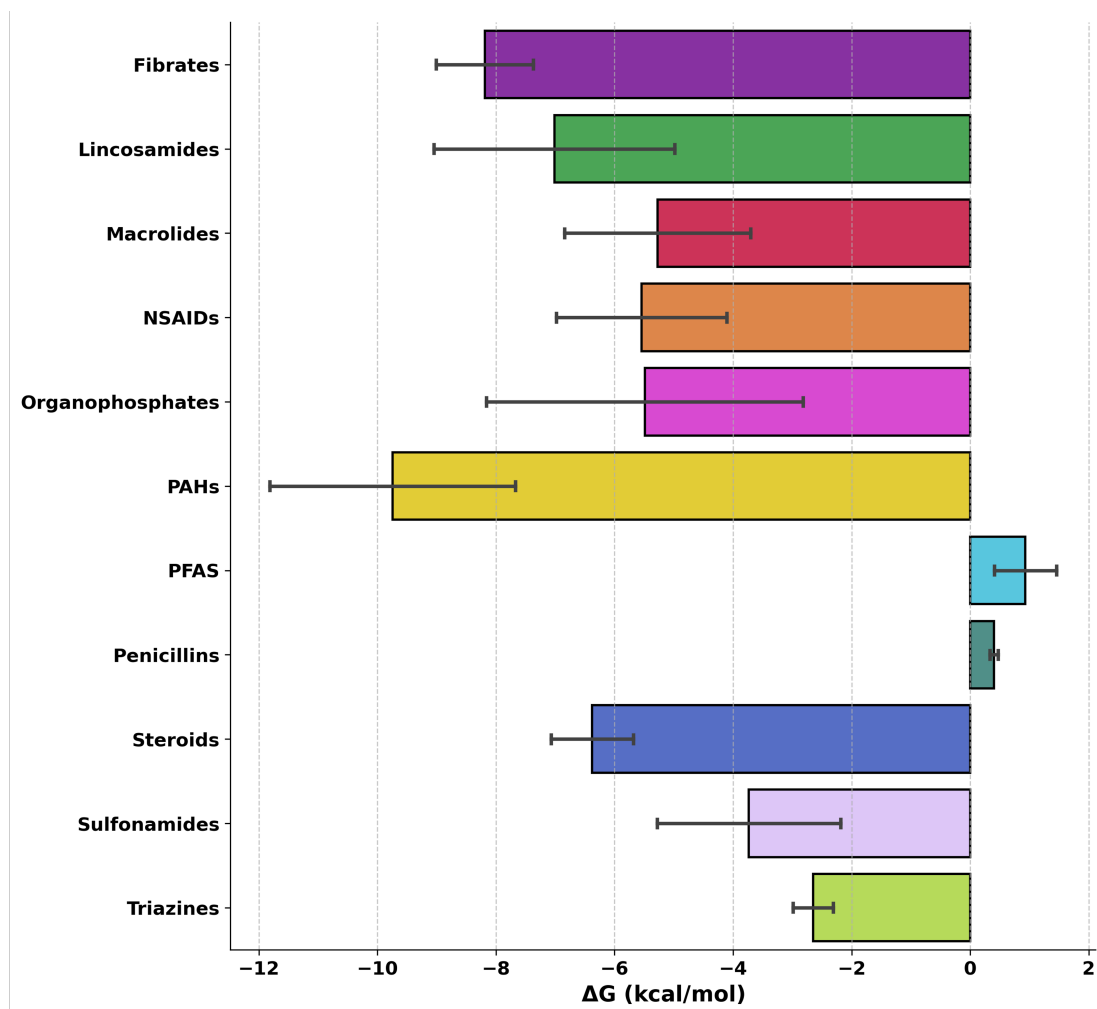


Figure 28: Average binding free energies across the identified chemical families on the GOX8-P935 surface. Values were computed using the NMODE-corrected MM-GBSA protocol (IGB = 5). Each bar represents the class average, while error bars denote the standard deviation within the cluster, reflecting the thermodynamic variance induced by different molecular scaffolds and heteroatomic patterns within the same family. Negative values indicate favorable adsorption, whereas positive values point toward electrostatic or steric repulsion. Classes are listed in alphabetical order.

As visualized in [Figure 28](#), the SAR analysis reveals disparities in the adsorption propensity of different chemical families on the highly oxygenated GOX8-P935 surface:

Polycyclic Aromatic Hydrocarbons (PAHs) and Fibrates: These classes represent the top-tier binders on the GOX surface. PAHs achieve the highest relative affinity (average -9.7 kcal/mol) through planar stacking interactions with the residual graphitic domains. Within this group, benzo-a-pyrene emerges as the most potent binder in the entire library (-12.1 kcal/mol), benefiting from its extended aromatic system. Fibrates follow closely (average -8.2 kcal/mol), demonstrating that even moderately sized aromatic systems can achieve stable adsorption when coupled with favorable van der Waals contacts.

Macrolide Antibiotics: Contrary to expectations based on size alone, the bulky macrolides show a much more moderate average affinity (-5.3 kcal/mol). Despite their high molecular weight (average 743.6 Da), their 3D complexity and sheer steric encumbrance hinder optimal contact with the highly functionalized GOX surface. The presence of dense oxygenated groups on the interface effectively creates a steric barrier that prevents these bulky structures from maximizing dispersive interactions, unlike the flat scaffolds of PAHs.

Organophosphate Insecticides: This class exhibits a moderate average binding affinity (-5.5 kcal/mol). Within this cluster, chlorpyrifos (-7.4 kcal/mol) outperforms diazinon (-3.6 kcal/mol), yielding a deviation from the cluster average of approximately ± 1.9 kcal/mol (Table 2). This internal variance is chemically justified by the topological differences between the two pesticides: while they share a thiophosphate core, the trichloropyridinyl ring of chlorpyrifos provides a more efficient stacking platform and additional halogen-mediated interactions compared to the isopropyl-pyrimidinyl scaffold of diazinon. This demonstrates how the clustering approach not only groups structurally related compounds but also provides a quantitative framework to pinpoint intra-class deviations.

Table 2: Intra-class thermodynamic variance within the Organophosphate Insecticides cluster. The table highlights the significant affinity gap between chlorpyrifos and diazinon on the GOX interface. The Δ from Avg metric quantifies how specific topological features, such as the trichloropyridinyl ring in chlorpyrifos, enhance binding stability compared to the class average.

Molecule	Charge	Chemical Class	Affinity (NMODE)	Class Avg	Δ from Avg
CHLORPYRIFOS	0	Organophosphates	-7.4	-5.5	-1.9
DIAZINON	0	Organophosphates	-3.6	-5.5	1.9

Perfluoroalkyl Substances (PFAS) and Penicillins: The clustering distinctly isolates highly electronegative and polar groups at the bottom of the affinity hierarchy. On the highly oxidized GOX surface, both penicillins and PFAS exhibit thermodynamically unfavorable, positive binding free energies (class averages of +0.4 and +0.9 kcal/mol, respectively). These results align with the previously discussed molecular insights: the dense array of surface carboxylates and epoxides exerts an overwhelming electrostatic repulsion against the anionic heads of PFAS (e.g., PFOS at +1.3 kcal/mol) and the highly polar, water-solvated structures of β -lactam antibiotics, effectively precluding stable adsorption.

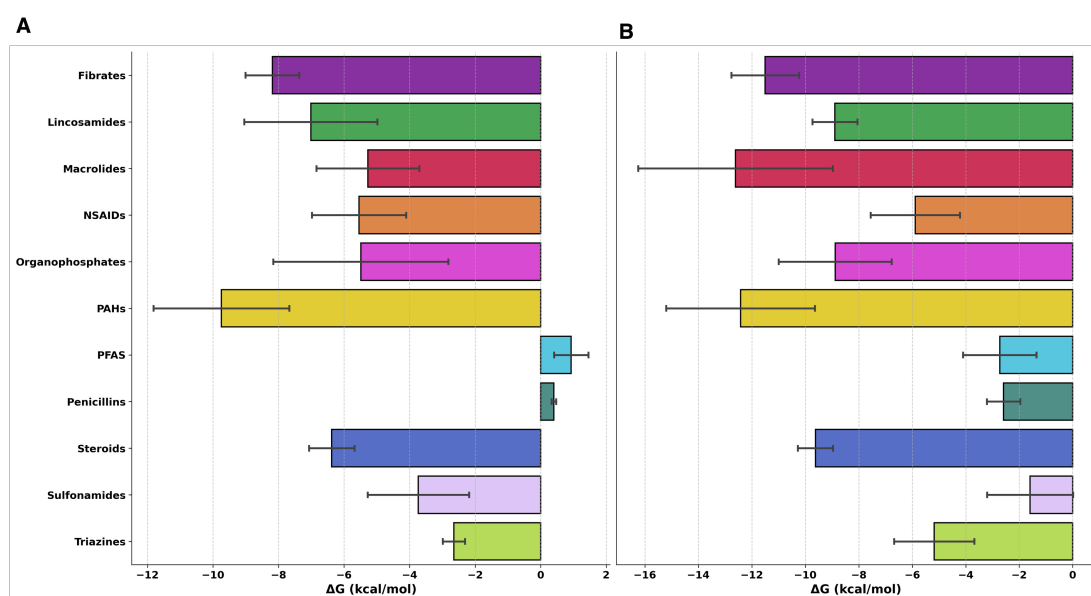


Figure 29: Comparative SAR analysis of the adsorption hierarchy on oxidized versus reduced graphene surfaces. Mean binding free energies are reported for the identified chemical classes on (A) the highly functionalized GOX8-P935 model and (B) the reduced REGOX8-P980 model. The side-by-side comparison captures the material’s transition from a selective, polar interface to a universal hydrophobic sponge. Notably, the restored planar topology of REGOX drives a dramatic affinity increase for bulky macrolides and enables the favorable adsorption of previously repelled species (PFAS and penicillins). Conversely, the depletion of surface oxygen sites penalizes classes reliant on specific hydrogen bonding, such as sulfonamides. Classes are listed in alphabetical order.

Crucially, comparing the clustered affinities reveals how the removal of oxygenated functional groups on the reduced surface (REGOX8-P980) not only enhances the overall affinity for neutral and hydrophobic classes by restoring the sp^2 carbon lattice, but it fundamentally alters the selectivity profile of the material. Consistent with the observations reported in the previous section (where the restored planar domains facilitated the optimal alignment of molecules like propranolol and azithromycin), the clustering data confirms this

trend on a macroscopic scale. As visualized in [Figure 29](#), the planar topology of REGOX allows bulky classes like macrolides – which were previously hindered by the corrugated GOX interface – to maximize dispersive interactions, resulting in a thermodynamic jump that places them at the top of the affinity ranking alongside PAHs and fibrates.

Furthermore, by neutralizing the repulsive electrostatic barrier, REGOX8-P980 becomes capable of binding previously rejected classes like PFAS and penicillins, as they transition from thermodynamically unfavorable, positive free energies on GOX to favorable, negative values on REGOX. Conversely, classes reliant on specific hydrogen bonding networks, such as sulfonamides, experience a significant drop in affinity on the reduced model due to the depletion of surface oxygen sites, demonstrating the transformation of a highly selective, polar interface (GOX) into a more universal, hydrophobic sponge (REGOX).

From an optimization perspective, this approach underscores that future computational screenings would greatly benefit from an expanded virtual library. Increasing the dataset size would naturally populate the current singletons, upgrading isolated data points into statistically robust clusters and further refining the predictive power of the SAR models for environmental remediation applications.

3.3.5 Protocol optimization

A final, essential evaluation of the computational protocol was performed to determine whether the binding affinity could be reliably estimated solely from the post-docking energy minimization step (ΔG_{min}), thereby obviating the need for the computationally intensive trajectory-based free energy analysis. The comparative results demonstrate, however, that ΔG_{min} does not constitute a robust or systematic predictor of the final binding free energy, as the discrepancy between the energy-minimized pose and the trajectory-averaged value exhibits substantial and non-systematic variability across distinct molecular scaffolds ([Figure 30](#)).

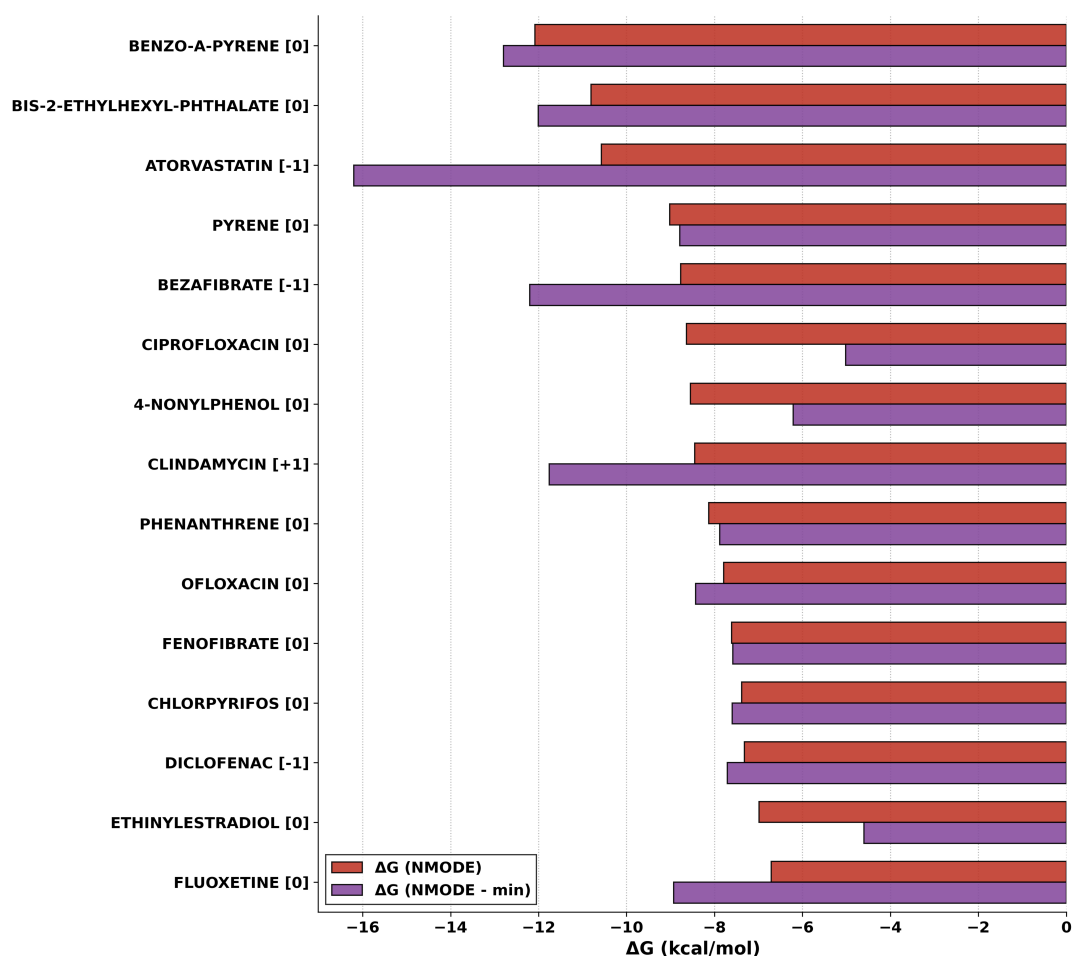


Figure 30: Comparison between trajectory-averaged and post-docking minimization binding affinities. The pronounced and irregular differences between the two methods across various compounds (e.g., the massive overestimation for atorvastatin compared to the stability of pyrene) highlight that energy-minimized poses alone are insufficient for a reliable thermodynamic ranking. The net charge of each pollutant is reported in brackets.

This discrepancy is quantitatively corroborated by analyzing the subset of poses that consistently attain the highest rankings: only 30% of the poses predicted by the IGB5 model coincide with the MD-derived energetic minima, with lower levels of agreement observed for the other implicit solvation models.

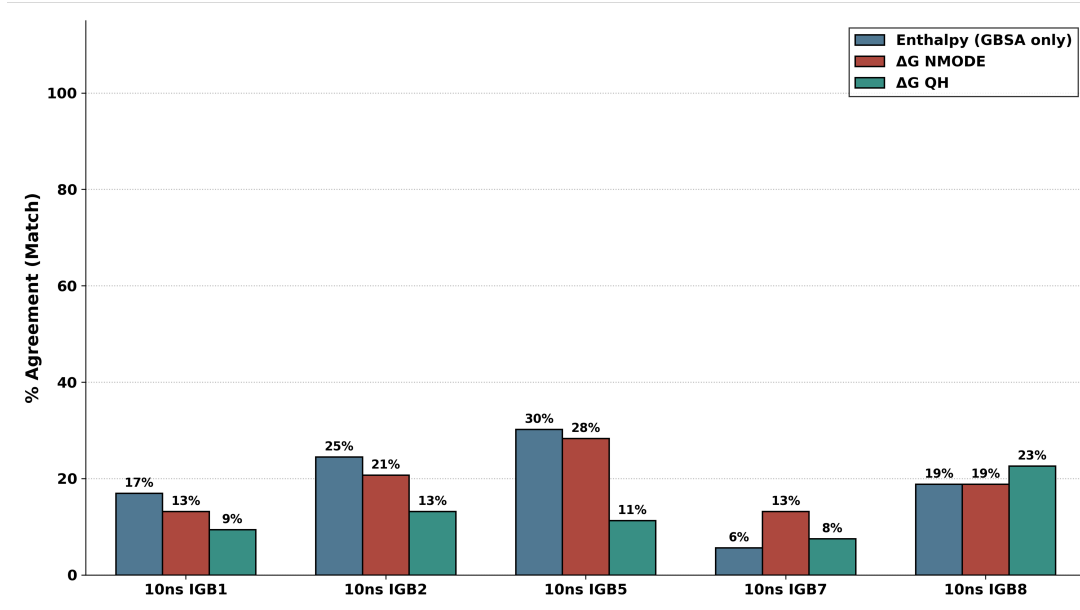


Figure 31: Pose agreement percentage across different GB solvation models. The maximum agreement is observed for the IGB5 model, demonstrating that the vast majority of minimized poses fail to coincide with trajectory-averaged minima across all tested methodologies.

This instability can be primarily ascribed to the intrinsically dynamic conformational landscape of the graphene oxide sheet, particularly the emergence of transient wrinkles and topological depressions (Figure 32) that are not captured by static energy minimizations performed on a rigid surface. Moreover, the temporal evolution of the interaction energy is highly ligand-dependent, with the global minimum occurring over a broad range of simulation times, spanning from early equilibration (e.g., frame 100) to the late stages of the trajectory (e.g., frame 800), indicating that each compound exhibits a distinct equilibration timescale before reaching its thermodynamically preferred adsorption configuration.

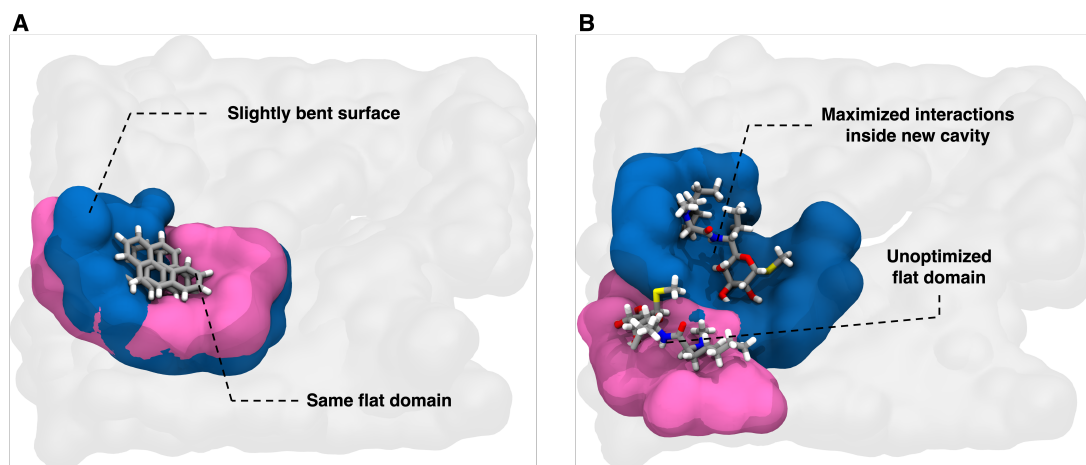


Figure 32: Impact of surface flexibility on ligand binding conformations. Surface representations illustrating the discrepancy between post-docking minimized poses (pink) and the minima obtained via AA-MD (blue). **(A)** The planar benzo(a)pyrene shows high spatial agreement between the two methods. **(B)** For clindamycin, the dynamic simulation allows the intrinsically flexible GO lattice to form transient wrinkles and topological depressions. This dynamically generated concavity better accommodates and engulfs the bulky ligand, explaining the difference in scoring such complexes.

Furthermore, the assessment of entropy evaluation methodologies revealed a pronounced dependence of the Quasi-Harmonic approximation on sampling density. A particularly illustrative case is bis(2-ethylhexyl) phthalate: whereas normal mode analysis yields an entropic contribution of approximately -17.0 kcal/mol, the QH approach considerably overestimates this term, giving a value of -40.2 kcal/mol when restricted to a 1000 frame trajectory. In contrast, increasing the sampling to 10000 frames produces a converged QH estimate of -18.1 kcal/mol, in close agreement with the NMODE reference. This behavior is consistent with established findings indicating that the QH approximation necessitates extensive sampling of the relevant phase space in order to yield reliable entropy estimates [240,241].

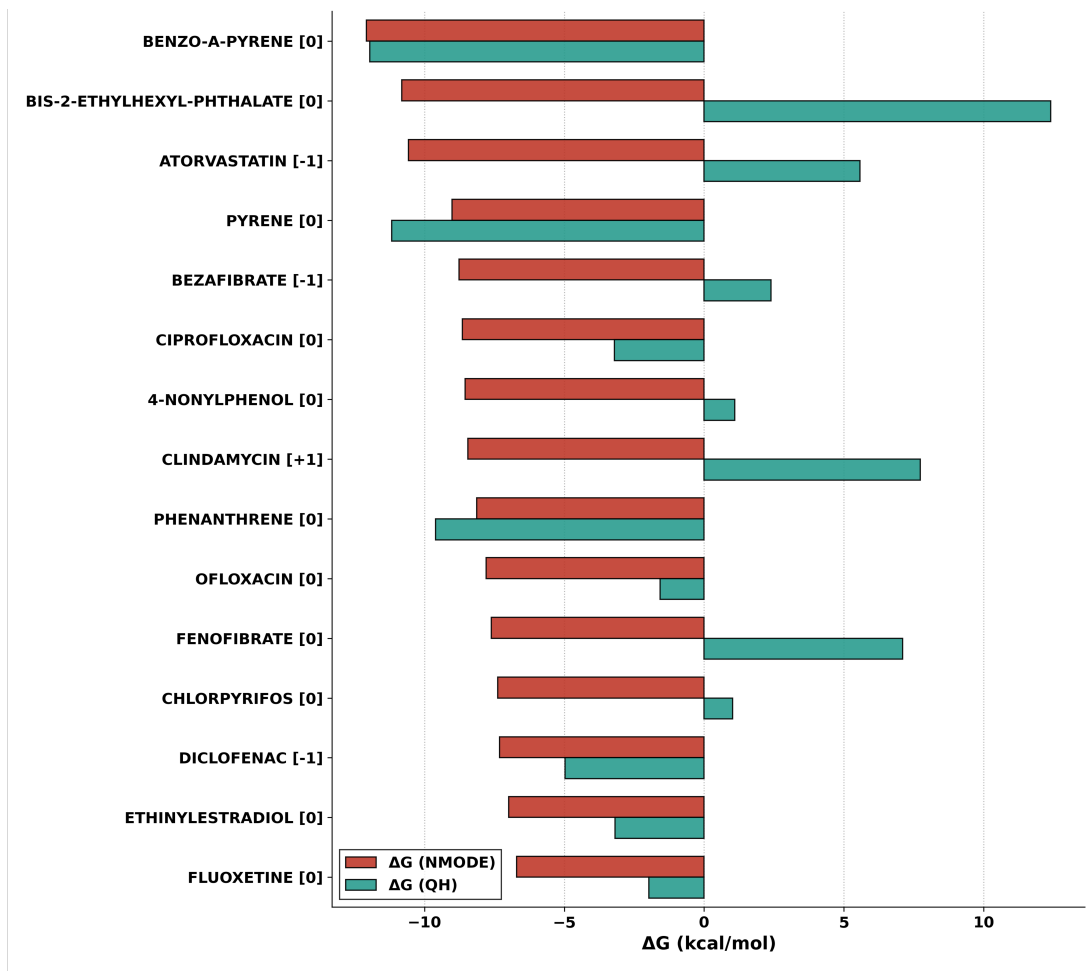


Figure 33: Impact of trajectory sampling density on Quasi-Harmonic entropy estimates. The noticeable overestimation of the entropic penalty for bis(2-ethylhexyl) phthalate illustrate that the QH method heavily depends on sampling density and requires extensive trajectory lengths to provide accurate thermodynamic contributions.

Consequently, future optimization efforts should investigate strategies such as docking ligands directly onto pre-equilibrated graphene oxide ensembles that exhibit fully developed surface features and systematically calibrating the duration of molecular dynamics simulations to ensure adequate sampling, with the main objective of validating the cost-effective quasi-harmonic method (≈ 10 minutes per trajectory) as a robust and reliable alternative to the computationally demanding normal mode analysis (≈ 7 hours per trajectory).

Moreover, to overcome the intrinsic timescale limitations of unbiased MD, the conformational space could be more comprehensively explored using enhanced sampling techniques such as Metadynamics [242] or Replica Exchange Molecular Dynamics (REMD) [243]. These approaches would enable the system to overcome high energy barriers and thereby facilitate the identification and

characterization of cryptic binding sites embedded within the dynamic folds and wrinkles of the graphene oxide sheet, which remain largely inaccessible to conventional MD simulations [\[244\]](#).

CHAPTER 4: COARSE-GRAINING

As demonstrated in the previous chapter, highly polar and anionic species such as PFAS (e.g., PFOA and PFOS) exhibit weak adsorption affinities on highly oxidized GO due to pronounced electrostatic repulsion. However, investigating their behavior solely as isolated monomers via atomistic MD severely limits the understanding of their true environmental dynamics, which are heavily dominated by concentration-dependent self-assembly and competitive adsorption in complex mixtures. To bridge this gap and access the extended timescales and system sizes necessary to simulate realistic aggregation and surface adsorption conditions, a coarse-grained approach was adopted.

In response to the growing recognition of the adverse human health and environmental impacts associated with per- and polyfluoroalkyl substances, a coarse-grained PFAS model compatible with the Martini 3 force field was developed in collaboration with the Groningen Molecular Dynamics group led by Prof. S. J. Marrink. In view of the large structural diversity and number of PFAS reported in the literature, 18 linear perfluorocarboxylic acids (PFCAs) and perfluorosulfonic acids (PFSAs) with varying chain lengths were selected for systematic parametrization.

Model validation was carried out following the established Martini protocol and was further extended by dedicated studies of self-aggregation behavior and interactions with graphene. Graphene was employed as a surrogate for the non-oxidized aromatic domains of graphene oxide, which are relevant for water remediation technologies, owing to the current lack of a validated Martini 3 model for graphene oxide. This approach provides a physically reasonable approximation of graphitic filtration regions.

The resulting parameter set displays excellent agreement with both all-atom molecular dynamics simulations and available experimental data, reliably reproducing key structural and physicochemical properties. These results establish a robust basis for the future development of automated parameter-generation workflows for PFAS within the Martini 3 framework. A detailed description of the model and its validation can be found in the original publication [\[245\]](#).

4.1 Methods

Despite the power of all-atom MD (AA-MD) simulations, they have been historically limited to tens of thousands of atoms and timescales of few microseconds, thus making the investigation of realistic complex systems a computationally heavy task. To overcome these limitations, coarse-grain (CG) approaches are used to reduce the resolution of molecular system to a simplified representation, grouping atoms into larger beads, thereby accelerating simulations due to the reduced number of degrees [246–252]. Within the realm of CG force fields, the Martini model is prevalently utilized in biophysics, biomedicine, nanotechnology, and materials design disciplines [253–261]. Notably, its latest incarnation, Martini 3, showcases enhanced interaction balance among particles of varying sizes and introduces numerous novel bead types, facilitating a more precise parametrization of diverse functional groups [262–264].

Martini 3 represents molecules with a discrete set of bead types that encode polarity/charge and come in three standard sizes: regular R, small S, and tiny T, corresponding to 4-to-1, 3-to-1 and 2-to-1 heavy-atom mappings, respectively. Neutral fragments fall into organic classes P (polar), N (intermediate), C (apolar) and X (halo-compounds), while charged fragments use Q (mono-valent) and D (divalent) beads and water uses its own W type. All bead categories are further subdivided by polarity and can be modified by labels for H-bonding (d/a), electronic polarization (e/v), charge character (p/n), self-interaction (h/r) and partial charges (q).

The non-bonded interactions are defined by a discretized Lennard-Jones interaction matrix with 22 levels and bead-size specific σ values, calibrated so that changing bead size or type reproduces systematic trends in hydrophobicity, miscibility and self-interaction [263].

Small-molecule parametrization in Martini 3 follows a building-block workflow: (1) a reference atomistic model is mapped to CG beads, respecting the default R/S/T sizes and avoiding splitting functional groups; (2) bead types are assigned by matching each fragment to reference entries or to bead partitioning data; (3) bonded parameters (bond lengths, angles, dihedrals) are fitted to atomistic simulations using center-of-geometry mapping including hydrogens and, if needed, refined to reproduce molecular volume, SASA and shape. For more complex or rigid aromatic systems, virtual interaction sites can be used to reduce constraint networks and keep simulations stable with a 20 fs

timesteps (default timestep in Martini), while still controlling ring torsions via tailored dihedrals. The small-molecule paper also provides a curated database of ~ 90 validated models, highlighting how to treat aliphatic and aromatic rings, substituents, and extended conjugated systems within this framework [265]. Depending on the investigated system, different strategies can be employed to validate the parametrization, including free energies of transfer between water and organic phases (e.g., hexadecane, octanol, chloroform), mixture miscibility trends, dimerization potentials of mean force (PMFs) for aromatic compounds, nucleobases, and transmembrane helices, lipid bilayer structural properties, and condensed phase densities.

The building-block philosophy, where beads represent generic chemical fragments with consistent interaction parameters across the force field, standardizes the parametrization procedure ensuring that new small molecules maintain compatibility with existing Martini components like lipids, proteins, and solvents. Such transferability enables multi-component simulations without ad hoc adjustments. The parametrization workflow is shown in Figure 34, as reported in the original Martini 3 paper.

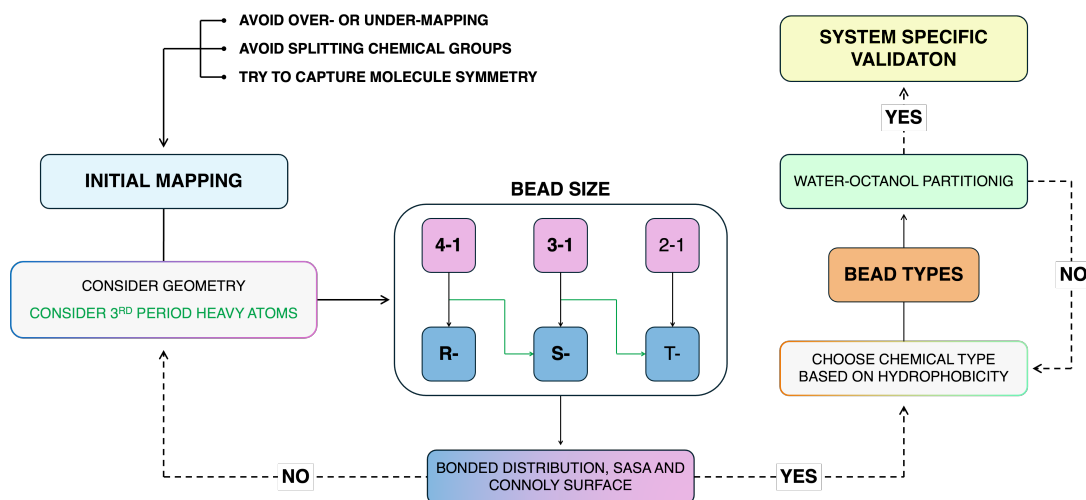


Figure 34: Coarse-grained parametrization protocol for Martini 3 models. Key steps include the initial geometric mapping, the determination of bead sizes based on heavy atom counts, and the selection of chemical bead types driven by hydrophobicity. The protocol relies on iterative refinement against structural targets and water-octanol partitioning data to achieve robust system-specific validation. Adapted from [263].

4.2 Martini 3 PFAS mapping and parametrization

All 18 PFAS considered in the present chapter were parametrized with the Martini 3 force field approach following the procedure outlined in Alessandri et al. [265] and the recommendations of the Martini 3 paper [263].

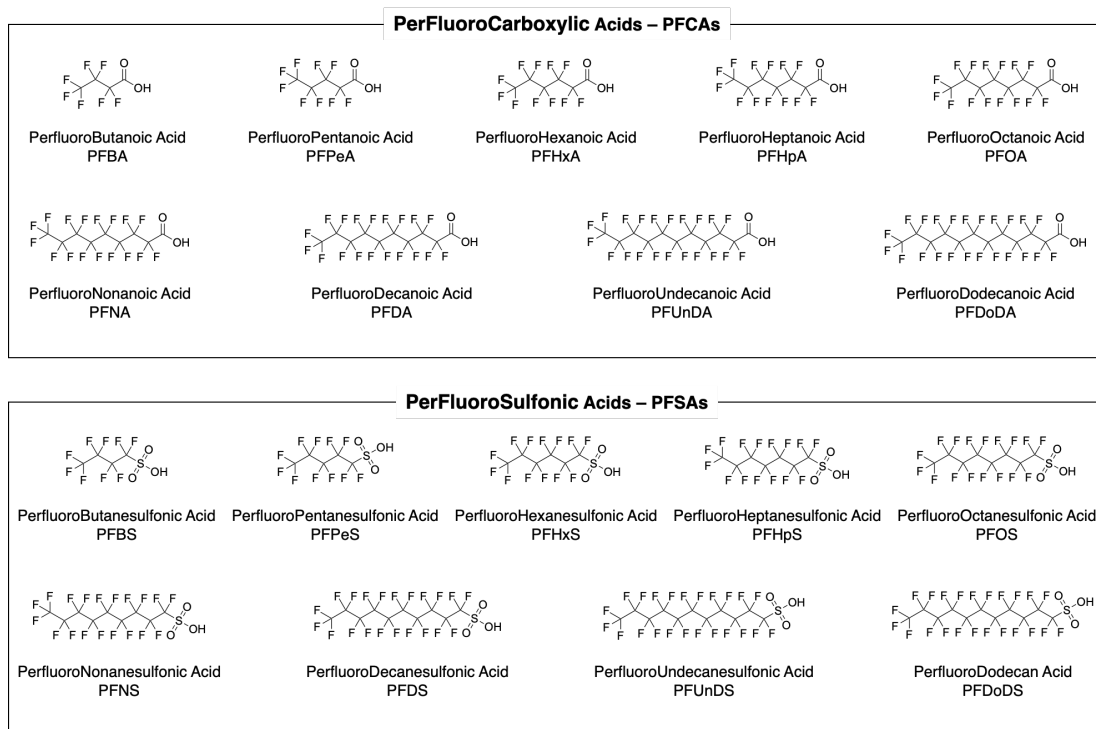


Figure 35: 2D representation of the PFAS considered in the present chapter. Nine PFCAs and nine PFSAs, with chain lengths from 4 to 12 carbons, were parametrized for Martini 3 force field.

The all-atom parameters of PFCAs and PFSAs with chain from 4 to 12 carbons were obtained following the standard AMBER parametrization using Gaussian 16 [106] and the AmberTools24 suite [100–102]. All molecules were considered ionized since they reportedly have a very low pKa [266–268] and the GAFF2 force field was used to model the PFAS. The AMBER topologies and coordinates were then converted to the GROMACS format using ParmEd v4.3.0 [269]. GROMACS 2023.4 [270–272] was used for the reference AA-MD simulations. Each molecule was solvated in a 5x5x5 nm³ cubic box with TIP3P water and underwent a 5000 step steepest-descent energy minimization. Proper system neutrality was ensured by adding Na⁺ counterions with *gmx genion*. Then, 10 ns NVT and NPT equilibrations were performed before going into full production, which consisted of a 500 ns simulation to ensure appropriate sampling of the bonded terms.

The equilibration and production steps were achieved via the md leap-frog integrator with a time step of 2 fs and the temperature was set to 300 K using the V-rescale thermostat with a tau-t of 1 ps. Pressure was maintained constant at 1 bar via C-rescale for the equilibration step and then with Parrinello-Rahman coupling for production. In both cases, isotropic coupling was used, while tau-p and compressibility were set to 5 ps and $4.5\text{e-}5\text{ bar}^{-1}$, respectively. To ensure consistency with the AMBER force field, van der Waals interactions were handled using a cutoff of 1.0 nm and the Potential-shift Verlet algorithm, while the long-range electrostatics were computed using Particle Mesh Ewald (PME) with a cutoff of 1 nm and a fourier grid spacing of 0.125 nm. Lastly, DispCorr was set to EnerPres and h-bonds were constrained using the LINCS algorithm.

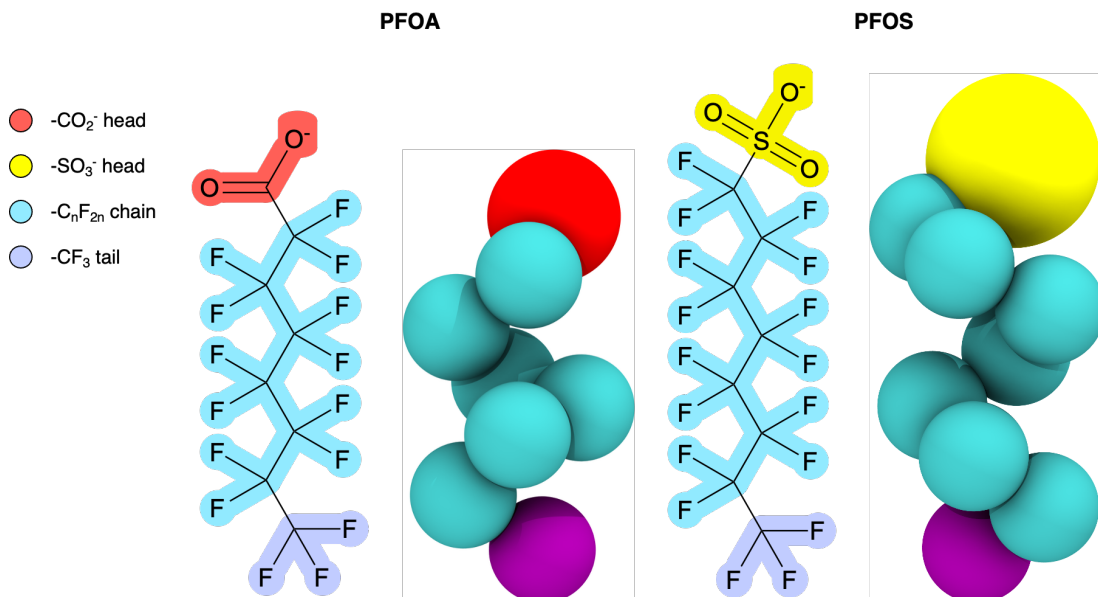


Figure 36: Breakdown of PFOA and PFOS into the three portions used for the CG mapping and their Martini 3 representation. In red, the carboxylate head was mapped to a SQ5n bead; in yellow, the sulfonate head was mapped to a Q3n bead; in cyan, the fluorine mainchain was mapped to TX3 beads; in purple, the trifluoromethyl tail was mapped to a SX4 bead.

To assign the bead types, the PFAS were divided into three portions: a head, a mainchain, and a tail (Figure 36). For PFCAs, an SQ5n bead was used to describe the carboxylate portion, TX3 beads were used for the middle $-\text{CF}_2$ moieties, and lastly a SX4 bead was chosen to represent the $-\text{CF}_3$ tail. The SQ5n bead was selected following the recommendation for carboxylate groups. The organofluorine groups were modeled as T-beads, considering one fewer non-hydrogen atom, in line with the description provided in the original Martini 3 paper [263], since fluorine atoms in halo compounds have small

atomic radii and exhibit weak van der Waals interactions. For PFASs, the middle and tail portions remained identical to those of PFCAs, while the headgroup moiety was mapped to a Q3n bead. Although the Martini 3 parametrization guidelines typically recommend a Q4n bead for sulfonate groups, the initial assessments indicated that the Q4n bead rendered the headgroup excessively polar, leading to significant deviations from experimental water/octanol partition coefficients. Consequently, the slightly less polar Q3n bead was selected as a data-driven optimization. This choice provided a substantially better agreement with experimental partition free energies, while still accurately capturing both the ionic nature and the higher bulkiness of the sulfonate group.

To parametrize the bonded parameters, CGBuilder [273] was used to map the reference AA trajectory, from which the target CG distributions were obtained. Great care was taken in representing the stiffness and helix conformation of the carbon chain, a key PFAS feature observed in the reference AA trajectory and widely reported in the literature [274–277]. In this Martini model, CG bonds and angles are described by harmonic potential energy functions (type 1 in GROMACS), while proper dihedral angles are obtained with a combination of periodic cosine dihedrals functions (type 9), as it allows multiple potential functions to be applied automatically to a single dihedral angle. The 1-3 and 1-4 interactions were excluded to ensure stability at 20 fs, in addition to the coupling between neighboring beads which is not considered by default. No improper dihedrals or virtual sites were used. Bond and angle parameters were kept consistent across all PFAS belonging to the same functional group category (carboxylates or sulfonates). Dihedral parameters were derived from those of PFOA or PFOS and subsequently refined on a case-by-case basis to improve system stability and parameter fittings.

Like the all-atom references, GROMACS 2023.4 [270–272] was used for the CG-MD simulations. CG PFCAs and PFASs were solvated in a 5x5x5 nm³ cubic box with regular water beads, neutralized with Na⁺ counterions, and underwent a series of energy minimization, 10 ns NVT and NPT equilibrations, and 500 ns NPT production. Bonds were kept relaxed during energy minimization and constrained at their equilibrium position throughout equilibration and production steps. The time step was set to 20 fs and the default Martini 3 GROMACS mdp parameters were used for all CG simulations, following the recommended settings [278]. Temperature was set to 300 K and managed with the V-rescale method, while pressure at 1 bar was

controlled by isotropic C-rescale (tau-p of 4 ps) during equilibration and Parrinello-Rahman (tau-p of 12 ps) coupling for the production. In both cases, the compressibility was set to $3\text{e-}4 \text{ bar}^{-1}$. Van der Waals interactions were handled using a cutoff of 1.0 nm and the Potential-shift Verlet algorithm, while the long-range electrostatics were computed using reaction-field with an epsilon_r of 15.

Comparison of the distributions (bonds, angles and dihedrals) between the final CG model and AA reference is shown in Figure 37 in case of PFOA. The data of all the other molecules can be found in the Appendix at the end of the present chapter. Overall, the AA reference data can be captured quite well. In some cases, in particular concerning the angles, fitting of multi-modal distributions resulted in system instability; in those cases, the parameters were modified to fit only the major peaks. Lastly, to preserve the characteristic stiffness and rigid geometry of the perfluorinated backbone [274–277], all bonds were converted into constraints. While employing flexible harmonic bonds with a reduced integration timestep (e.g., $\leq 10 \text{ fs}$) was considered, it would have required exceptionally high force constants to counteract the multiple dihedral potentials, leading to persistent numerical instabilities. Thus, using constraints proved to be the most robust approach to prevent the artificial floppiness of the chains, maintaining the structural integrity of the PFAS models while preserving the standard 20 fs timestep.

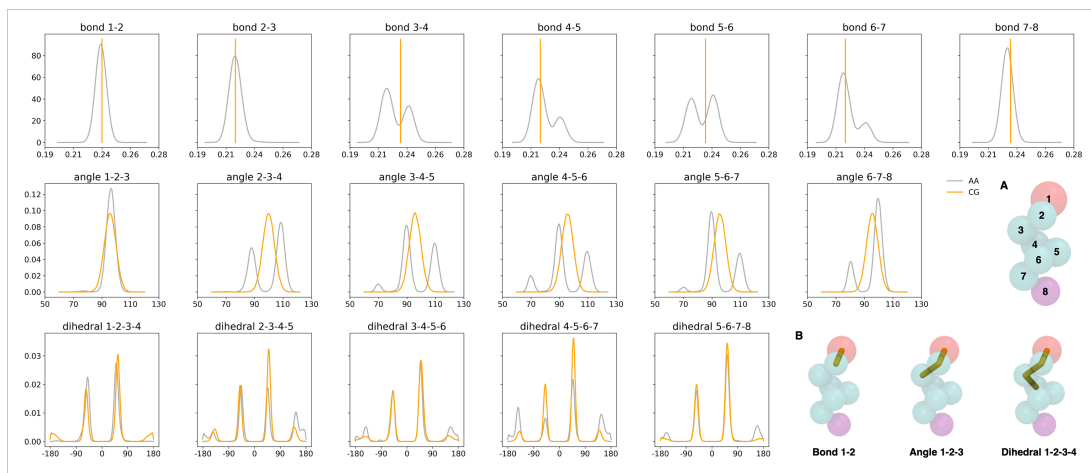


Figure 37: Comparison of the bonded distributions between the mapped AA reference and the final CG trajectory in case of PFOA. For the bonds, the CG distribution is represented as a single peak since they were frozen throughout the simulations. On the right side, (A) the numbering of the beads in the CG representation; (B) highlight of first bond, angle and dihedral.

4.3 Model validation

4.3.1 SASA and free energies of transfer

To validate the created topologies, the solvent accessible surface area (SASA) and the Connolly surface of the AA and CG models were compared. The SASA and Connolly surfaces were computed with the *gmx sasa* command. For the AA reference, the updated VdW radii were taken from Rowland and Taylor [279], while the standard radii of Martini 3 R-, S- and T- beads were used for the CG trajectories. The SASA is slightly underestimated in the Martini 3 model, but still within acceptable error range (Appendix, Figure 63). The Connolly surfaces do not show large discrepancies, suggesting the overall shape is represented well (Figure 38).

In addition, octanol/water partitioning free energies were computed to obtain the logP values. However, since these molecules are ionized and exhibit surfactant behavior, it is more appropriate to refer to logD. This naming inconsistency and the wide variability in both experimental and predicted logP/logD values have been emphasized by Forest and Rayne [280,281], who attributed these discrepancies primarily to the challenges inherent in measuring partitioning properties for perfluorinated surfactants using common techniques. These difficulties stem from the fact that PFAS tend to form three immiscible layers when they are added to an octanol/water mixture because they are simultaneously hydrophobic and lipophobic, leading to surfactant aggregation, interfacial adsorption, and ionization equilibria influencing the effective partitioning behavior [282,283]. To ensure the reliability and coherence of the parametrization, the experimental logD determined via cyclic voltammetry by Jing et al. [284] were used as reference, as this data is already used to model the partitioning, environmental fate and toxicological mechanisms of anionic perfluoroalkyl carboxylic and sulfonic acids [285,286]. To ensure comprehensive validation against both experimental and computational reference data, the parametrization was additionally benchmarked against quantitative structure-activity relationship (QSAR) predictions reported by Torralba-Sanchez et al. [286], where the authors expanded Jing’s experimental dataset predicting anionic partition coefficients from neutral species.

Free energies of transfer from water to octanol were computed by umbrella sampling (US) simulations performed with the CG model. In US simulations,

the molecule was placed inside an equilibrated biphasic $6 \times 6 \times 13$ nm³ box composed of water and hydrated octanol. Umbrella windows were spaced 0.1 nm apart along the distance between the center of mass (COM) of the PFAS and the water phase. Each window was simulated for 500 ns and the distance was kept fixed by an umbrella potential with a force constant of 1500 kJ mol⁻¹ nm⁻². The energy profiles were calculated using the weighted histogram analysis method (WHAM) 60 as implemented in the GROMACS tool *gmx wham*. Semi-isotropic Parrinello-Rahman was used for pressure coupling. The compressibility and tau-p were set to 3×10^{-4} bar⁻¹ and 1 ps on the z axis, and 0 on the xy plane. Other mdp parameters were kept the same as the previous CG simulations.

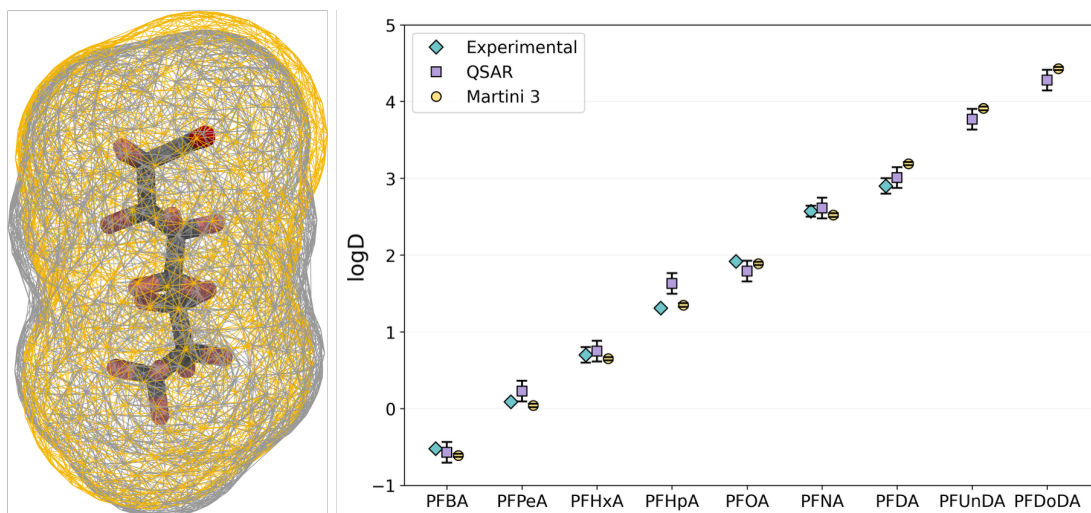


Figure 38: Comparison of PFOA Connolly surface and PFCAs logD values. On the left, the comparison of the Connolly surface between the AA (grey) and CG (orange) model; on the right, comparison of logD values between Martini 3 (circle), experimental (diamond) and QSAR (square) predicted data. When available, error data is reported. The hydrophobicity trend, and therefore the surfactant behavior, is preserved at the CG level.

The comparison with reference logD values of PFCAs is shown in Figure 38. logD values exhibits a very high agreement with both experimental and QSAR predicted data, validating the choice of bead types. Similarly to the PFCAs, the computed logD values for PFSA show very good consistency with the limited experimental evidence available – namely the PFBS and PFOS values from Jing et al. – and with the expected chain-length dependence, despite the general lack of data for this subclass of compounds (Appendix, Figure 66).

4.3.2 Adsorption on graphene

PFAS molecules, due to their amphiphilic nature, are known to adsorb spontaneously on carbon-based materials like graphene [287,288], driven primarily by hydrophobic and van der Waals interactions as well as electrostatic forces depending on surface functionalization and environmental conditions [289,290].

Given the environmental importance of PFOA as a representative PFAS molecule [291,292], the model was further validated by reproducing its adsorption on pristine graphene. This choice is motivated by the analysis performed by Wang et al. [293], which provides detailed insights into the energetics and mechanisms of PFOA adsorption on different graphene surfaces, demonstrating a spontaneous and energetically favorable binding, driven mainly by hydrophobic interactions between the perfluorocarbon tail and the graphene surface.

A periodic sheet of Martini 3 graphene was created with the code provided in Shreshta et al. [259] and solvated in a 10x10x20 nm³ cubic box with a 60:40 ratio of Regular and Tiny water. This approach was adopted to reduce the unwanted and unrealistic layering effect of Regular Martini water described by Cambiaso et al. [294]. Periodic Boundary Conditions (PBC) were applied in all three dimensions. The graphene sheet was built to be continuous across the x and y boundaries, effectively creating an infinite 2D layer. Importantly, the graphene beads were left completely unrestrained, allowing the sheet to naturally exhibit conformational fluctuations and rippling, albeit to a reduced extent due to the CG resolution. Along the z-axis, the 20 nm box length provided a sufficiently large solvent buffer to allow the complete desorption of the PFOA molecule into the bulk phase during the umbrella sampling, thus avoiding any spurious interactions with its periodic images. Then, the same umbrella sampling procedure described above was applied to follow the desorption along the distance between the graphene and PFOA center of mass. Simulations were conducted at 300 K and compared to the reference AA model obtained with the OPLS-AA force field [295]. Anisotropic pressure coupling was adopted, with compressibilities of 3e-4 bar⁻¹ on the diagonal and 0 on the off-diagonal.

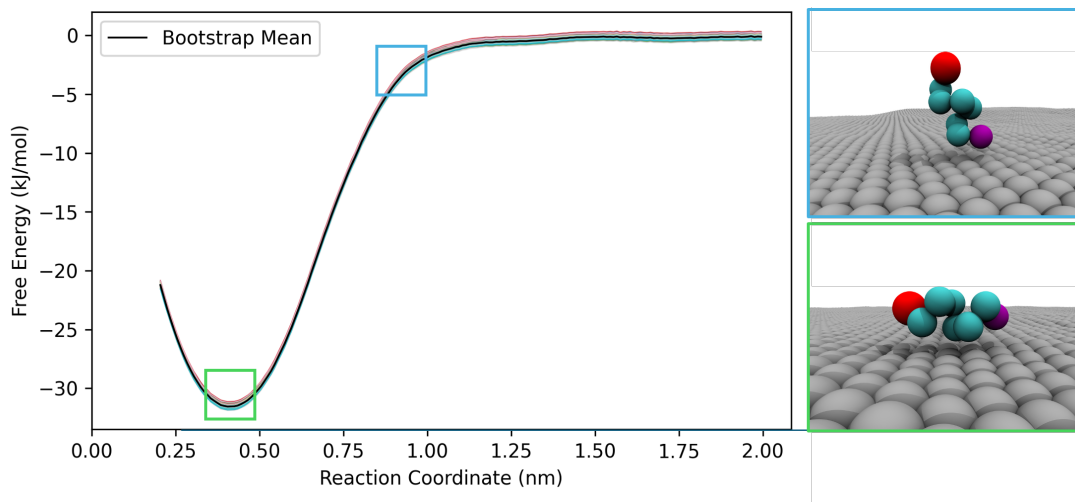


Figure 39: PMF of PFOA desorption from graphene. The conformations present at the adsorption minima (green) and before complete desorption (blue) are similar to the ones observed in previous OPLS-AA simulations [295].

As shown in Figure 39, the PFOA was pulled from its starting, adsorbed state on the graphene surface towards the bulk of the solvent. A clear minimum corresponding to an energy well of about -31.2 ± 0.2 kJ/mol is present at ~ 0.45 nm, where the hydrophobic interactions between the fluorinated beads and the carbon surface are maximized. At ~ 0.9 nm, only the tail portion is seen interacting with graphene before complete desorption. These results agree very well with OPLS-AA data, both in terms of reported ΔG (-29.9 kJ/mol) and PFAS conformations, therefore confirming the validity and parametrization of Martini 3 linear PFAS and graphene.

To further validate the choice of bead type, the same procedure was repeated by adding the *e* electron-rich label to PFOA beads in a twofold manner: 1) following the molecular electrostatic potential map (MEP), therefore treating only the α -carbon as electron-rich; 2) considering all the fluorinated chain beads as electron-rich. The MEP was computed using Gaussian16 [106] at the PBE0/6-311++G* level of theory and basis set. The results (Appendix, Figure 69) show a stronger adsorption to the graphene surface as more *e* labels are added, further deviating from the reference data, although the MEP-based approach still gives a reasonable PMF. Because of this, the unlabeled X-type beads were maintained.

4.3.3 Phase behavior

A peculiar aspect of PFAS molecules is their ability to form supramolecular aggregates such as micelles. To assess the Martini parametrization, systems ranging from 5 mM to 40 mM of PFOA were simulated in a 30x30x30 nm³ water box for 2 μ s to ensure system stability and sample the equilibrated phase. The simulation setup was maintained identical to the previous CG dynamics, and the *gmx clustsize* command alongside SciPy [296] were utilized for data analysis and visualization. Initial baseline simulations included only neutralizing counterions to establish overall charge neutrality, whereas subsequent models incorporated 0.25 M NaCl to specifically investigate salting-out effects and environmental ionic conditions.

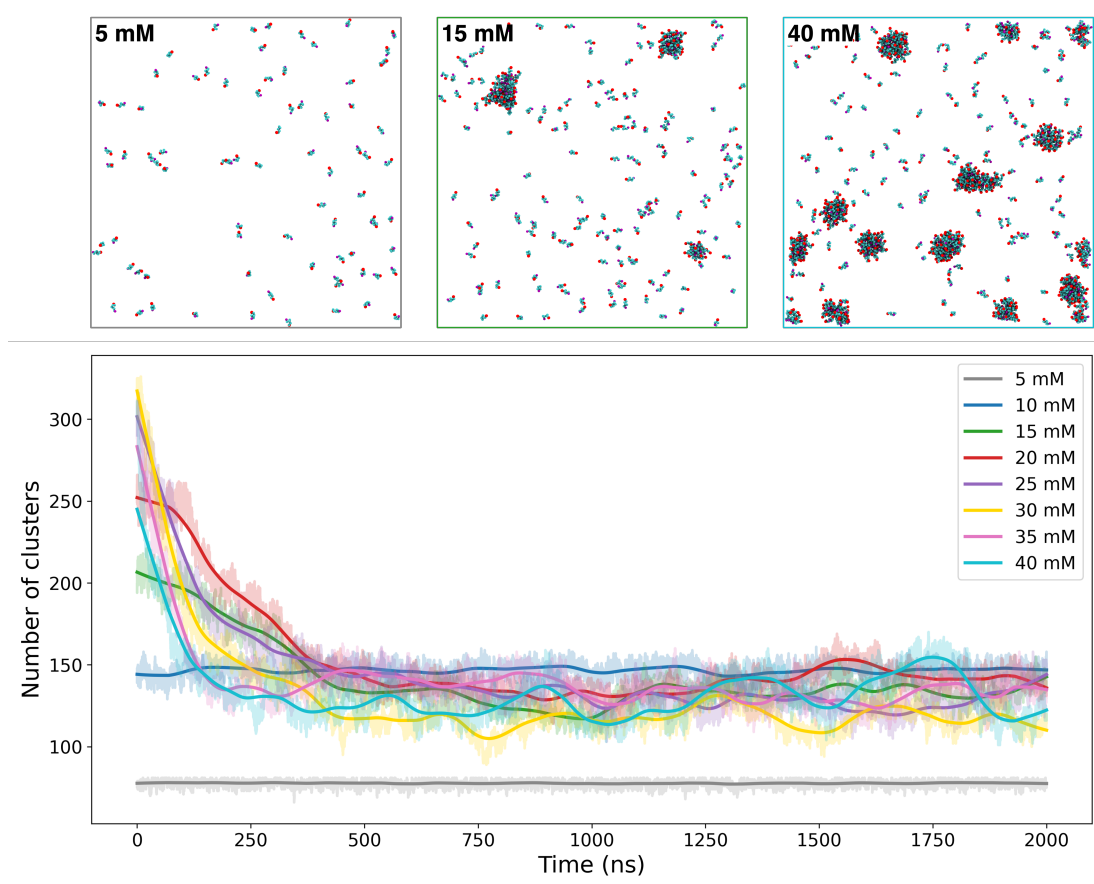


Figure 40: The number of clusters at increasing PFOA concentrations. The micelles begin forming at ~ 15 mM and become more frequent with increasing concentrations, as observed in the representative frames above. Water and ions molecules were removed for visual clarity.

Figure 40 illustrates the number of clusters, which are defined as any group of molecules that are associated on a distance-based criterion, observed over the course of the simulations. At 5 and 10 mM, the number of clusters remains

approximately constant throughout the simulation, indicating that the system predominantly consists of monomeric surfactants and transient small aggregates, with no stable structures appearing. Starting from 15 mM, a pronounced decrease in the number of clusters is observed, becoming greater with increasing PFOA concentration. This reflects the formation of larger, more stable aggregates (i.e. micelles), which reduces the total number of discrete clusters in the system. Between 15 mM and 25 mM the curves reach a plateau toward the end of the simulation, signaling the onset of an equilibrium phase characterized by a dynamic exchange of monomers between micelles and the bulk solution.

To quantify the dynamic exchange of monomers between the micellar core and the bulk solution, a dual-metric analysis was implemented using *in-house* Python scripts based on the MDAnalysis [297] library. The thermodynamic residence time (τ) was calculated from the state autocorrelation function $C(t)$, where each monomer i was assigned a binary state variable $S_i(t)$: $S_i = 1$ if the monomer's distance from the micellar center of mass was below $R_{in} = 1.0 \text{ nm}$ (first minimum of the monomer-monomer RDF), and $S_i = 0$ otherwise. The brackets $\langle \cdot \rangle$ denote ensemble and time averages over all valid monomers. The long-time tail ($t > 10 \text{ ns}$) of $C(t)$ was fit to a single exponential decay $C(t) = e^{-t/\tau}$ to extract the mean residence time, excluding the initial fast relaxation regime associated with interfacial rattling [298,299]. The autocorrelation function of the state fluctuations was computed as:

$$C(t) = \frac{\sum_i \langle S_i(0) S_i(t) \rangle}{\sum_i \langle S_i^2 \rangle} \quad (38)$$

The macroscopic exchange rate was determined using a spatial double-cutoff criterion to isolate complete solvation events from transient corona fluctuations. A desorption event was counted when a monomer's distance from the micellar center of mass irreversibly transitioned from $R_{in} = 1.0 \text{ nm}$ (micellar core boundary) to $R_{out} = 3.0 \text{ nm}$ (bulk limit). Uncertainties associated with the results were quantified using a block-averaging protocol, in which the 1 μs equilibrated trajectory was partitioned into ten independent blocks. The final reported values represent the mean and the standard deviation across these portions.

Above $\geq 20 \text{ mM}$, the macroscopic exchange rate remains constant at ~ 250 fully solvated monomers μs^{-1} micelle $^{-1}$. While faster than experimental

fluorosurfactant rates [300], these nanosecond/microsecond kinetics are characteristic of Martini CG models, where reduced resolution smoothens the free-energy landscape, lowering desorption barriers while preserving equilibrium structures and thermodynamics [261,263]. At higher concentrations (30 mM onwards), a more oscillating time evolution is observed due to the increased interaction of the different clusters present in solution, leading to transient formation of larger clusters. This finding is also confirmed by the polydispersity observed in the size distribution plots (Appendix, Figure 70). Specifically, the average aggregation number (N_{agg}) increases from ~ 39 at 20 mM to ~ 50 at 40 mM, indicating a morphological transition from ideal spherical micelles to more elongated structures. Interestingly, the thermodynamic residence time τ of individual monomers decreases from 59 ns (20 mM) to 40 ns (40 mM), consistent with Aniansson-Wall stepwise kinetics [301]. Despite this faster desorption accompanying the morphological transition, the macroscopic exchange rate remains constant, demonstrating robust steady-state flux independent of aggregate morphology.

At higher concentrations (30 mM onwards), a more oscillating time evolution is observed due to the increased interaction of the different clusters present in solution, leading to transient formation of larger clusters. This finding is also confirmed by the polydispersity observed in the size distribution plots. Based on the obtained data, the critical micelle concentration (CMC) can be placed around ~ 15 -25 mM, which is in line with models obtained with other CG force fields present in literature [302], albeit somewhat lower than the expected experimental CMC of NaPFO, which ranges between 25 and 38 mM [303–311]. The exchange kinetics across the micellar regime is reported in Table 3:

Table 3: The average aggregation number (N_{agg}), the mean residence time of a surfactant monomer within the micellar aggregate (τ), and the resulting monomer exchange rate ($\mu\text{s}^{-1} \text{ micelle}^{-1}$) for PFOA systems at different concentrations. All values are reported as mean $\pm$ standard deviation calculated by using a block-averaging protocol, in which the 1 μs equilibrated trajectory was partitioned into ten independent blocks. A significantly higher standard deviation is observed at the lowest concentration, attributed to the proximity of the system to the CMC.

Concentration (mM)	Tau (ns)	N_{agg}	Exchange rate
15	62	34	289 ± 204
20	59	39	255 ± 28
25	53	41	246 ± 54
30	45	45	250 ± 45
35	45	48	248 ± 25
40	40	50	249 ± 19

To see if the model could be improved in this regard, the influence of additional bead labels was investigated (Figure 41). In line with the results for adsorption on graphene, addition of e or r labels did not improve the aggregation behavior, typically resulting in a complete lack of micelle formation.

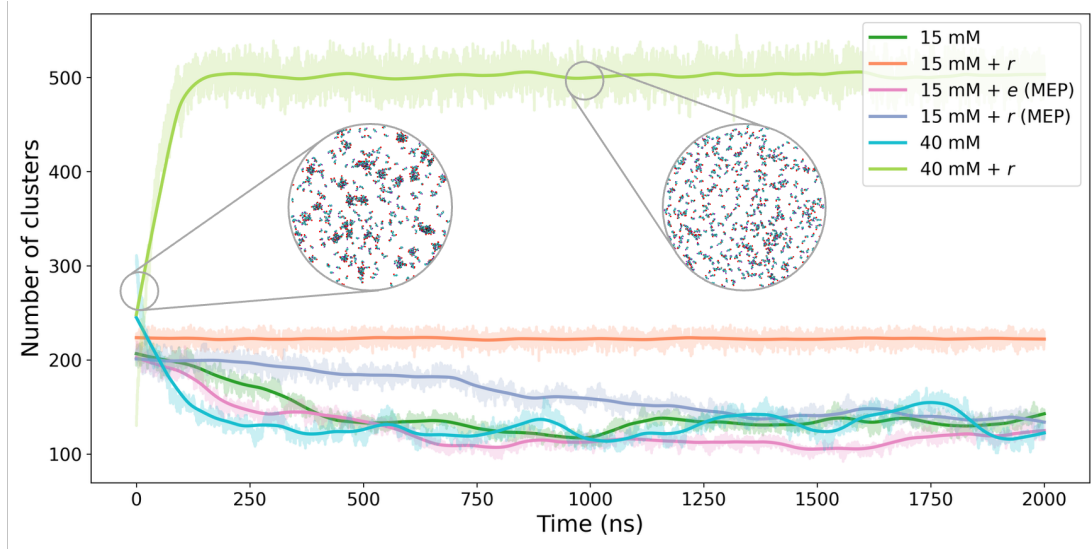


Figure 41: The number of clusters at increasing PFOA concentrations with labeled e and r beads. Both labels slow the kinetics of micelle formation and completely impede the aggregation at higher concentrations. Note that, unlike the other analysis, the starting frame for the 40 mM consisted in an already aggregated state that quickly falls apart and reverts to a completely monomeric solution. Water and ions were removed for visual clarity.

In addition, the effect of increasing the salt concentration was tested. According to experimental data [304,308], PFOA aggregation is facilitated at

higher salt concentrations by a salting out effect. As shown in Figure 42, the PFOA model can reproduce this behavior, at least qualitatively, showing formation of micelles at lower surfactant concentrations in the presence of 0.25 M NaCl.

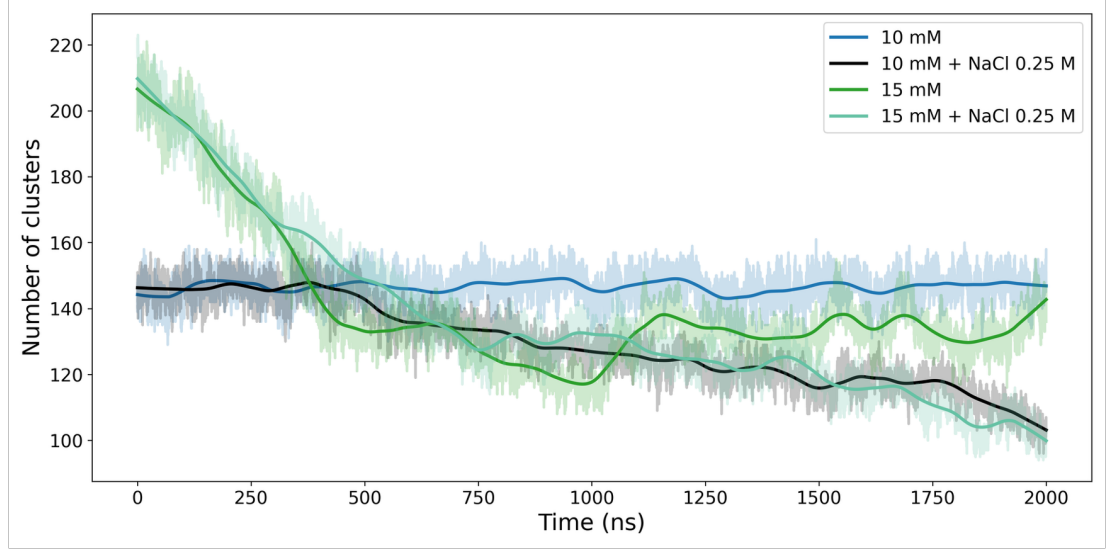


Figure 42: The number of clusters at increasing PFOA concentrations with the addition of NaCl 0.25 M. The increased NaCl concentration lowers the effective CMC due to a salting out effect. Water and ions were removed for visual clarity.

To further validate the electrostatics of the CG model the zeta-potential (ζ) was estimated from the equilibrium MD trajectories through a semi-analytical model of the electrical double layer. For each simulated concentration, the structural micellar radius was extracted from the spatial distribution of the terminal polar SQ5n beads (SQ5n – SQ5n). The thickness of the Stern layer and the fraction of strongly associated Na^+ counterions were obtained from the first minimum of the SQ5n – Na^+ radial distribution function. Because the micellar radius is of the same order of magnitude as the Debye length (κ^{-1}), planar approximations such as the Grahame equation may overestimate the surface potential. Therefore, the finite curvature of the nanometer-sized micelles was accounted for by computing the surface potential (Ψ_0) from the effective surface charge density (σ) using the Ohshima analytical approximation for charged spherical colloids [312]:

$$\sigma = \frac{2\varepsilon_r\varepsilon_0\kappa k_B T}{ze} \sinh\left(\frac{ze\Psi_0}{2k_B T}\right) \left[1 + \frac{2}{\kappa R_{surf} \cosh^2(ze\Psi_0/4k_B T)}\right]^{1/2} \quad (39)$$

The macroscopic zeta-potential (ζ) was then obtained by propagating Ψ_0 to the hydrodynamic slipping plane (R_{slip}) using a spherical Poisson-Boltzmann decay [313]:

$$\zeta = \Psi_0 \left(\frac{R_{surf}}{R_{slip}} \right) e^{-\kappa(R_{slip}-R_{surf})} \quad (40)$$

$$\kappa = \sqrt{\frac{2N_A e^2 c_0}{\varepsilon_r \varepsilon_0 k_B T}} \quad (41)$$

To avoid arbitrary assumptions about the hydration-shell thickness, the slipping-plane (R_{slip}) position was defined as the sum of the Stern-layer thickness (δ_{Stern}) and the Lennard-Jones collision diameter (σ_{LJ}) of the corresponding Martini 3 solvent beads [263].

$$R_{slip} = R + \delta_{Stern} + \sigma_{LJ} \quad (42)$$

Once again, uncertainties associated with the results were quantified using a block-averaging protocol, in which the 1 μ s equilibrated trajectory was partitioned into ten independent blocks. The final reported values represent the mean and the standard deviation across these portions.

Table 4 shows how model captures colloidal charge-regulation behavior [314,315]: as the total solute concentration increases from 15 to 40 mM, the fraction of condensed sodium increases from $11.7 \pm 1.1\%$ to $36.6 \pm 0.9\%$ thereby reducing electrostatic repulsion. Consequently, the calculated ζ -potentials across the different water models and concentrations fall approximately in the -49 to -72 mV range. Notably, at 40 mM, the models yields ζ -potentials ranging from -49.2 ± 0.3 mV (Regular water) to -56.3 ± 0.3 mV (Tiny water), in good agreement with the experimental reference value of -59 ± 10 mV for NaPFO micelles [316].

Table 4: The parameters used in estimating the ζ -potential. The slipping plane position was defined by adding the Lennard-Jones collision diameters of the different Martini 3 water models (Regular, Small, and Tiny) to the Stern layer boundary. As the concentration increases from 15 to 40 mM, the fraction of condensed counterions rises, effectively buffering the increase in N_{agg} . All values are reported as mean $\pm$ standard deviation calculated by using a block-averaging protocol, in which the 1 μ s equilibrated trajectory was partitioned into ten independent blocks. Similar to the kinetic data, the 15 mM system exhibits slightly higher standard deviations, reflecting larger equilibrium fluctuations in micellar size and shape as the system approaches the CMC.

Concentration (mM)	N_{agg}	Micelle radius (nm)	Stern layer (nm)	Condensed ions (%)	Free ions (%)	Effective charge	Ψ_0 (mV)	Zeta potential (mV) R water	Zeta potential (mV) S water	Zeta potential (mV) T water
15	34	0.95 ± 0.05	0.58 ± 0.01	11.7 ± 1.1	88.3 ± 1.1	-30.0 ± 0.4	-203.1 ± 6.4	-63.5 ± 0.8	-67.0 ± 0.9	-71.5 ± 1.0
20	39	1.05 ± 0.03	0.58 ± 0.01	20.5 ± 0.8	79.5 ± 0.8	-31.0 ± 0.3	-187.9 ± 2.9	-59.2 ± 0.7	-62.6 ± 0.7	-66.8 ± 0.8
25	41	1.07 ± 0.03	0.58 ± 0.01	25.5 ± 1.1	74.5 ± 1.1	-30.6 ± 0.4	-180.9 ± 3.8	-55.4 ± 0.5	-58.7 ± 0.6	-62.8 ± 0.7
30	45	1.07 ± 0.02	0.58 ± 0.01	29.0 ± 0.8	71.0 ± 0.8	-32.0 ± 0.4	-179.7 ± 2.4	-53.2 ± 0.4	-56.5 ± 0.4	-60.5 ± 0.5
35	48	1.08 ± 0.02	0.58 ± 0.01	33.3 ± 1.0	66.7 ± 1.0	-32.0 ± 0.5	-176.6 ± 2.3	-51.4 ± 0.3	-54.6 ± 0.3	-58.7 ± 0.3
40	50	1.08 ± 0.01	0.58 ± 0.01	36.6 ± 0.9	63.4 ± 0.9	-31.7 ± 0.5	-174.1 ± 1.4	-49.2 ± 0.3	-52.4 ± 0.3	-56.3 ± 0.3

Nonetheless, whereas PFOA is expected to form micelles, longer chain PFAS can form lamellar structures owing to the higher hydrophobicity and rigidity of the perfluorinated chain which can impact the critical packing parameter (CPP) [317–319].

Beyond the intrinsic chain length, the supramolecular assembly of PFAS is also highly sensitive to the specific electrostatic environment. While simple monovalent counterions (e.g., Na^+) typically favor micellization for shorter chains, the introduction of multivalent cations or oppositely charged cationic surfactants can actively crosslink or neutralize the anionic headgroups. This drastically alters the CPP, driving the spontaneous formation of stable thin bilayers and vesicles, as demonstrated in previous experimental and computational efforts on catanionic PFAS mixtures [320,321]. Furthermore, such ion- and co-surfactant-mediated interactions are known to modulate the aggregation thermodynamics, significantly lowering the CMC [316].

To explore this intrinsic tendency for higher-order assembly in this CG model, two larger systems encompassing ~ 1600 PFCAs of different length, namely PFOA (8 Cs) and PFDoDA (12 Cs), both at a concentration of ~ 100 mM – way above the CMC – and NaCl 0.15 M were simulated for 20 μ s.

Despite starting from an identical condition, the systems evolve into different structures: PFOA aggregates to form slightly elongated micelles, while PFDoDA completely aggregates into a single spherical assembly. The temporal evolution of PFOA and PFDoDA aggregation is presented in Figure 43.

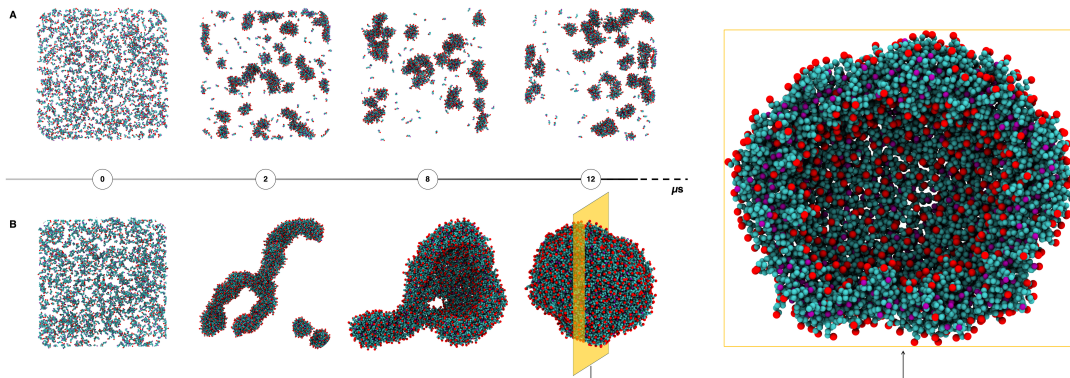


Figure 43: Temporal evolution of PFOA and PFDODA aggregates. (A) The monomers expectedly aggregate to form micelles that do not evolve to more complex structures. (B) The solution quickly transitions from a monomeric solution to elongated micelles and worm-like structures. After 8 μs , the assembly starts to bend and folds to yield a unilamellar vesicle. Although not perfectly spherical, the vesicle remains structurally stable over the final microseconds of the simulation. The cross section of the PFDODA spherical vesicle-like aggregate reveals an arrangement similar to that of a bilayer. Water and ions molecules were removed for visual clarity.

In the latter case, the initial solution of monomers rapidly self-assembles into elongated worm-like micellar structures of approximately 2 μs . At 8 μs , all aggregates have coalesced into a single supramolecular assembly that progressively acquires curvature. This curved structure then folds onto itself to yield a unilamellar vesicle from 12 μs onwards. Although the vesicle does not attain a perfectly spherical morphology, it remains structurally stable over the final microseconds of the simulation. By visually investigating the cross section, the overall arrangement resembles that of a bilayer, with the head beads (red) exposed to the solvent, the tail (cyan) beads pointing inwards, and the terminal tail beads (purple) distributed around the bilayer midplane.

4.4 Complex pollutants mixture simulations

To demonstrate the capabilities and applicability of the present model, two case studies were qualitatively investigated. The first consisted in a mixture of 7 PFAS molecules differing in head group and perfluoroalkyl chain length, designed to emulate a mature industrial leachate [322–327]. The second mimicked a mixture comprising 13 emerging contaminants, containing next to PFAS a number of small molecules, selected on the basis of recent literature [109,328–336], the availability in the Martini Database (MAD) [337] and by cross-referencing our previously developed emerging contaminants database (EC-DB). The exact composition of both mixtures is presented in Table 5.

Table 5: Composition of the pollutants mixtures system. In all cases, the systems had a total contaminant concentration of 100 mM and were simulated in presence of NaCl 0.15 M in a 30x30x30 nm³ cubic box. The relative concentrations assigned to the various non-PFAS micropollutants in the model mixture were not chosen to replicate any particular field site. Instead, they were selected to mirror the commonly reported patterns of occurrence and concentration for emerging contaminants in landfill leachate and urban wastewater: highest levels for persistent solvents such as 1,4-dioxane and phenolic compounds; intermediate levels for volatile aromatics (e.g., benzene) and industrial aromatics (e.g., aniline); and lower levels for heterocyclic tracers including benzothiazole, indole, caffeine, and p-cymene.

Contaminant	N° of molecules	Contaminant	N° of molecules
PFBA	407	PFOA	488
PFPeA	325	PFOS	325
PFHxA	209	1,4-Dioxane	163
PFOA	209	Phenol	163
PFNA	68	Benzene	122
PFBS	130	Aniline	98
PFHxS	139	1,2,4-Trichlorobenzene	65
PFOS	139	1,2-Dichlorobenzene	65
		Anthracene	65
		Benzothiazole	41
		Indole	16
		Caffeine	8
		p-cymene	8

Each system was prepared at a total solute concentration of 100 mM in the presence of 0.15 M NaCl and solvated in a 30x30x30 nm³ water box. Following an initial equilibration phase of 10 ns, performed in line with prior CG simulations, production runs were carried out for 10 μ s at 300 K.

The same two systems were subsequently replicated in the presence of a periodic Martini 3 graphene slab and a mixed-solvent environment composed of Regular and Tiny water in a 60:40 ratio in order to characterize their aggregation properties and adsorption behavior. Given the current lack of a validated graphene oxide model with the Martini 3 force field, graphene was employed as a surrogate for the non-oxidized aromatic domains of GO used for water-remediation [338,339], thereby providing a reasonable approximation of

graphitic filtration regions. For the systems without graphene, isotropic pressure coupling was employed. For the systems containing graphene, anisotropic pressure coupling was used, with compressibilities set to $3 \times 10^{-4} \text{ bar}^{-1}$ along the diagonal elements of the pressure tensor and to 0 for the off-diagonal elements.

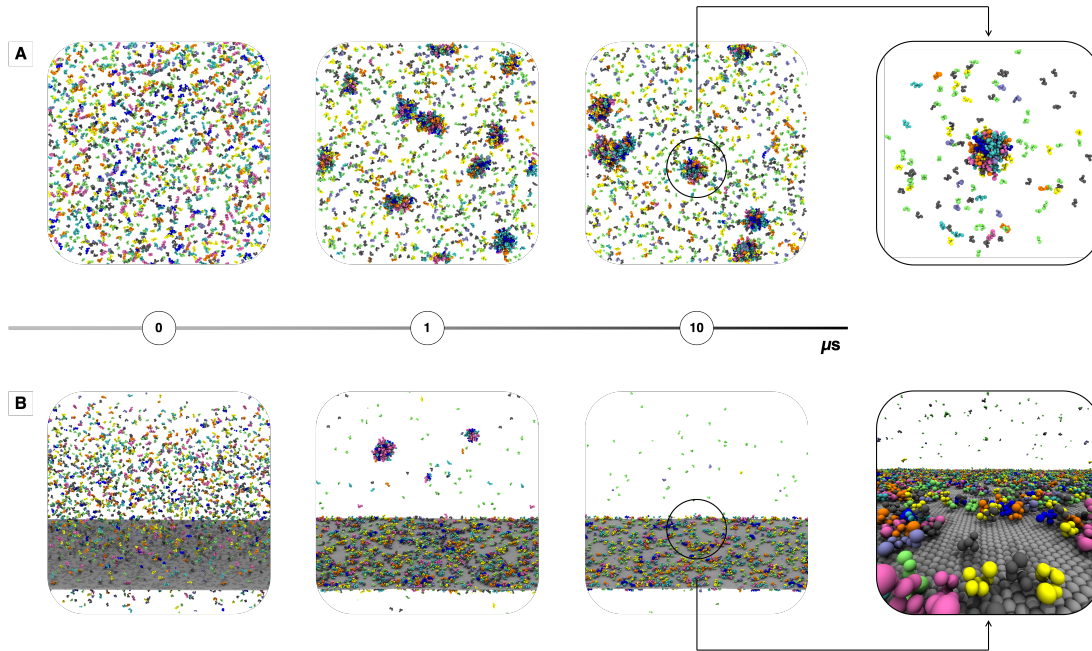


Figure 44: Temporal evolution of the PFAS mixture in the absence and presence of graphene. (A) More hydrophobic PFAS tend to self-assemble into mixed micelles, whereas more hydrophilic ones remain as monomers in solution. (B) The addition of graphene suppresses micelle formation. The different PFAS are colored as follows: PFOA in cyan; PFOS in mauve; PFBA in green; PFPeA in gray; PFHxA in yellow; PFNA in blue; PFBS in light blue; PFHxS in orange. Water and ions were removed for visual clarity.

For the PFAS-only mixture, the observed trends are consistent with those obtained in the previous simulations. Longer-chain, more hydrophobic PFAS (PFOA, PFNA, PFOS and PFHxS) tend to self-assemble into mixed micelles, whereas shorter-chain, more hydrophilic PFAS (PFBA, PFPeA, PFHxA and PFBS) exhibit transient interactions with the micelle surfaces but mostly remain as monomers in solution. Upon addition of graphene, almost all PFAS, except PFBA, are adsorbed onto the graphene surface, ultimately suppressing micelle formation. The corresponding time evolution for the PFAS mixture is reported [Figure 44](#).

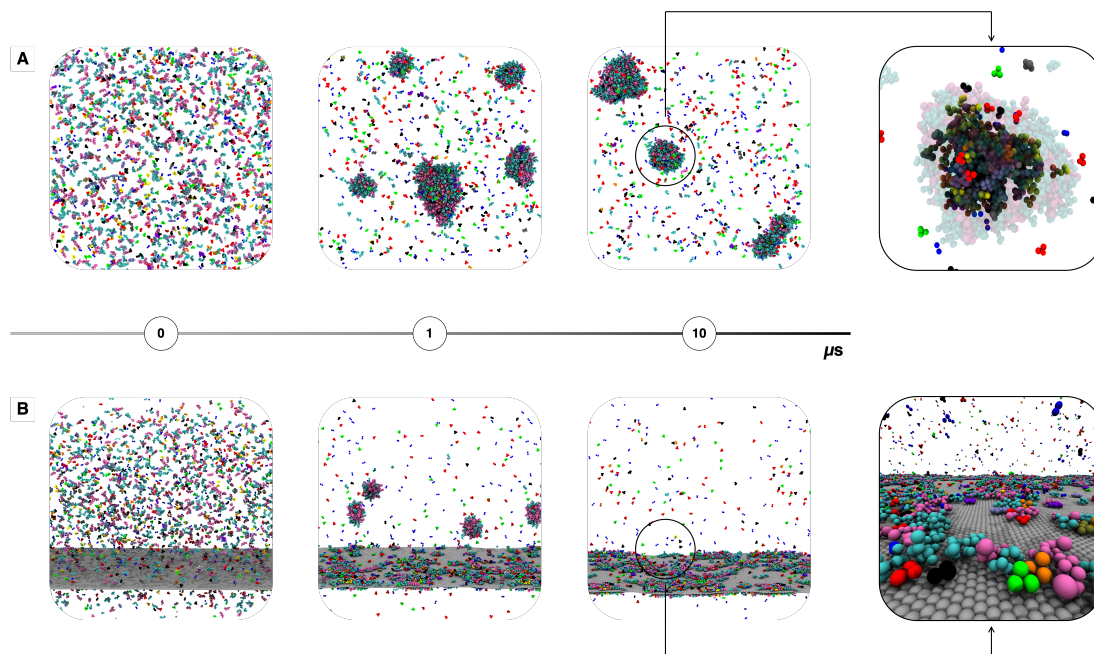


Figure 45: Temporal evolution of the emerging contaminants mixture in the absence and presence of graphene. (A) PFAS molecules embed hydrophobic contaminants inside micelle-like aggregates. **(B)** Pollutants adsorb on the graphene surface forming flat aggregate domains. The different pollutants are colored as follows: PFOA in cyan; PFOS in mauve; 1,4-dioxane in blue; phenol in red; benzene in black; aniline in green; 1,2,4-trichlorobenzene in yellow; 1,2-dichlorobenzene in orange; anthracene in light blue; benzothiazole in gray; indole in ochre; caffeine in violet, p-cymene in white. Water and ions molecules were removed for visual clarity.

Figure 45 illustrates the temporal evolution of the system with the mixture of emerging contaminants. At first inspection, PFOA and PFOS appear to form micellar aggregates within approximately 1 μs , whereas several other pollutants (1,4-dioxane, phenol, aniline) predominantly remain dispersed in the aqueous phase. A more detailed analysis of one of these micellar aggregates, however, reveals that the PFAS molecules primarily constitute an external shell surrounding a strongly hydrophobic core composed mainly of 1,2,4-trichlorobenzene, 1,2-dichlorobenzene, benzothiazole, benzene, and anthracene. This behavior is particularly relevant because it can significantly influence the fate and transport of these compounds, potentially modifying their overall biodistribution and toxicity as well as the efficiency and mechanisms of their remediation.

The introduction of graphene induces a distinct modification in the behavior of the system: smaller, layered micellar structures persist, although to a lesser extent, as most pollutants become adsorbed onto the graphene surface. At 10 μs , nearly all pollutants are adsorbed onto the graphene, forming predominantly flat aggregate domains. However, because the simulation setup exposes both sides of the graphene sheet to the solvent, the available adsorption

area is effectively doubled. Therefore, although the number of simulated molecules is insufficient to fully saturate this expanded surface, it is hypothesized that upon complete occupation of the primary layer, a secondary, pollutant-mediated adsorption layer would develop. This secondary layer could either impede or promote the subsequent adsorption of additional contaminants.

Although these results provide a primarily qualitative description of the studied systems, they clearly demonstrate the capability of this PFAS model to reliably capture the behavior of environmentally relevant contaminant mixtures.

4.5 Appendix

Presented herein are the additional visual data and validation metrics complementary to the study reported in Chapter 4.2. This material mirrors the Supporting Information (SI) of the associated publication [245] and is included herein to ensure the comprehensive documentation of the computational protocol.

Table 6: All the considered system in modelling the phase behavior. In all cases, the systems were neutralized with Na^+ and enclosed in a $30 \times 30 \times 30 \text{ nm}^3$ cubic box.

Concentration (mM)	N° of PFOA molecules	Concentration (mM)	N° of PFOA molecules
5	81	10 + NaCl 0.25 M	162
10	162	15 + NaCl 0.25 M	243
15	243	15 + e label	243
20	324	15 + e label (MEP)	243
25	406	15 + r label	243
30	487	15 + r label (MEP)	243
35	568	40 + r label	649
40	649	Concentration (mM)	N° of PFDODA molecules
100 + NaCl 0.15 M	1626	100 + NaCl 0.15 M	1626

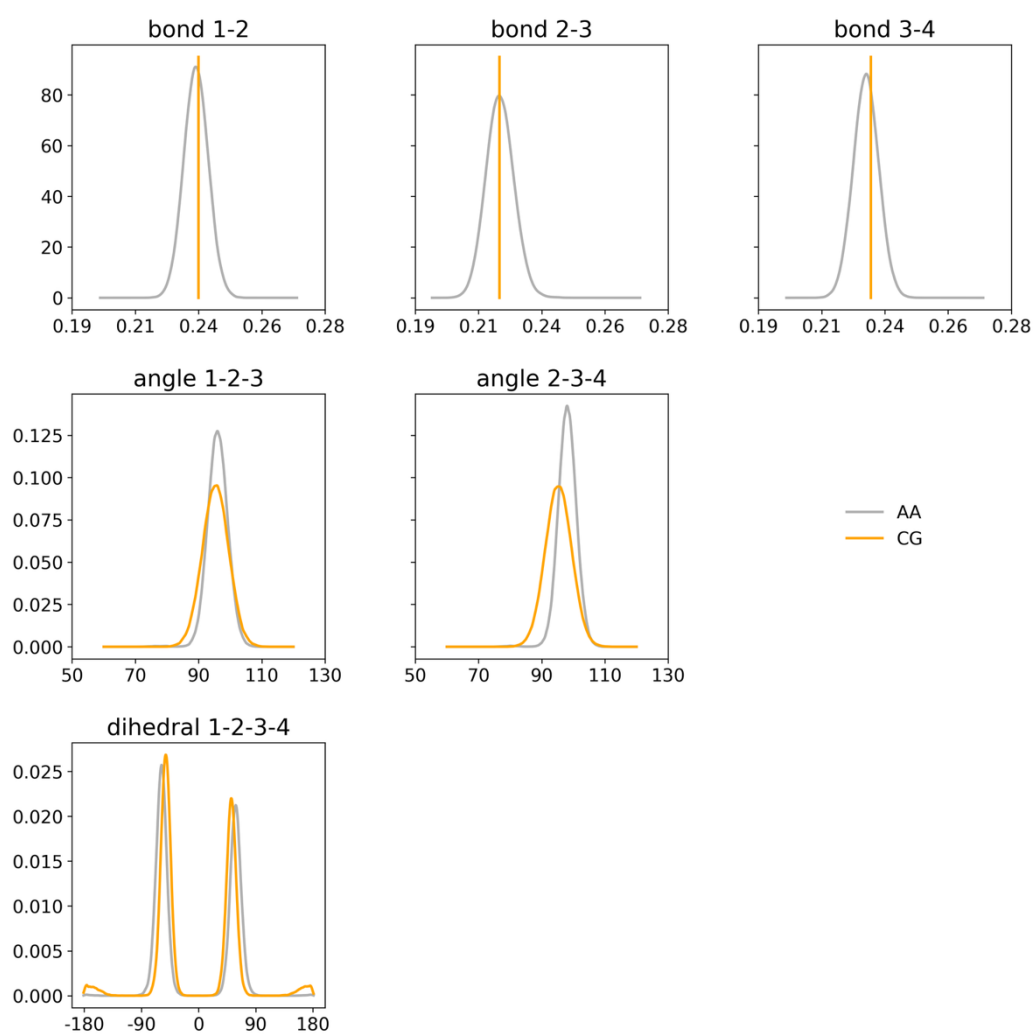


Figure 46: The bonded distribution of PFBA (perfluorobutanoic acid).

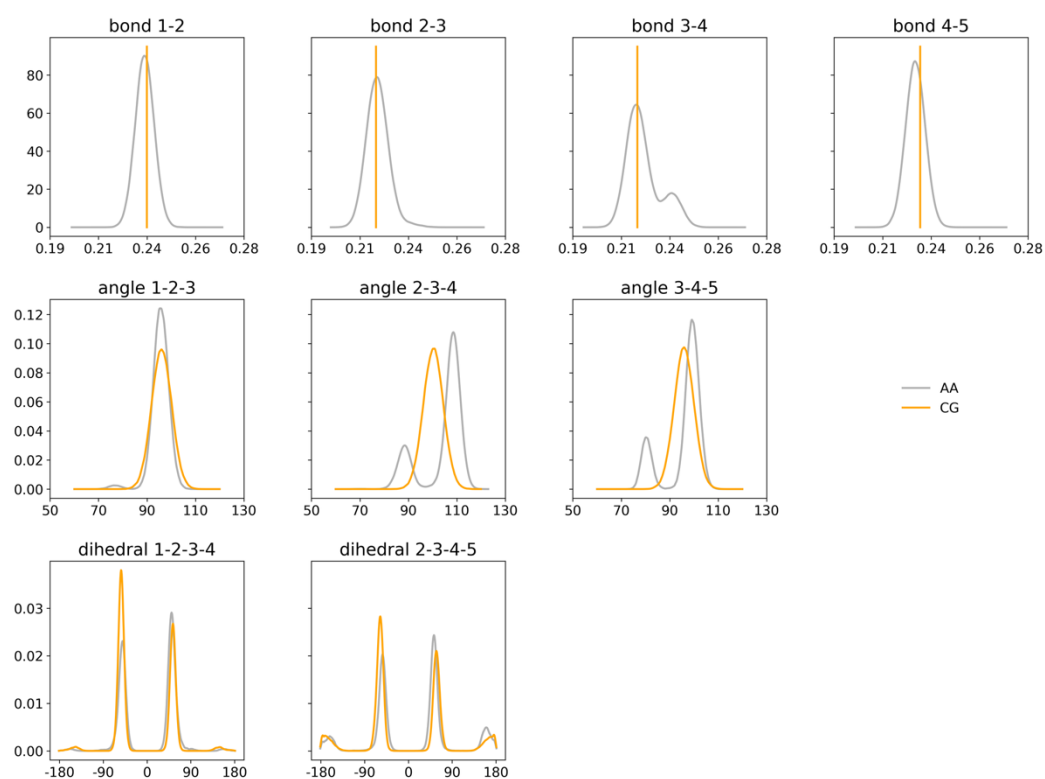


Figure 47: The bonded distribution of PFPeA (perfluoropentanoic acid).

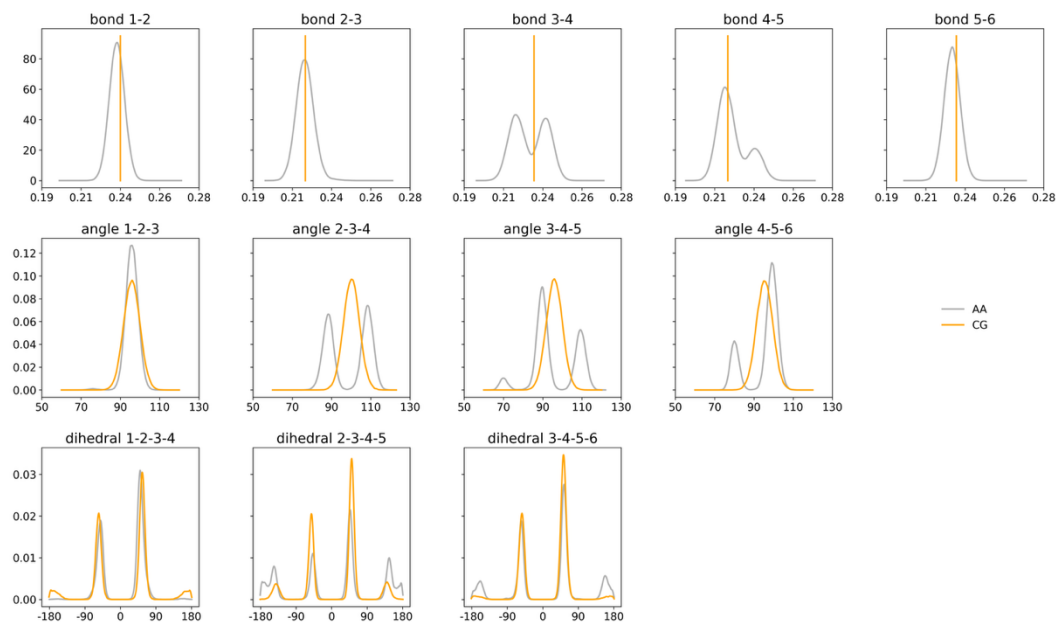


Figure 48: The bonded distribution of PFHxA (perfluorohexanoic acid).

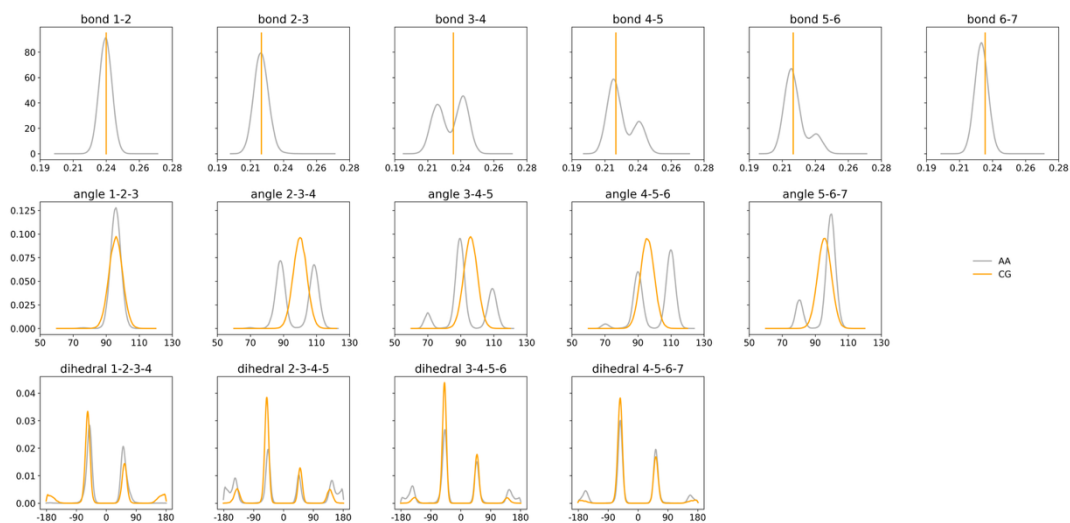


Figure 49: The bonded distribution of PFHpA (perfluoroheptanoic acid).

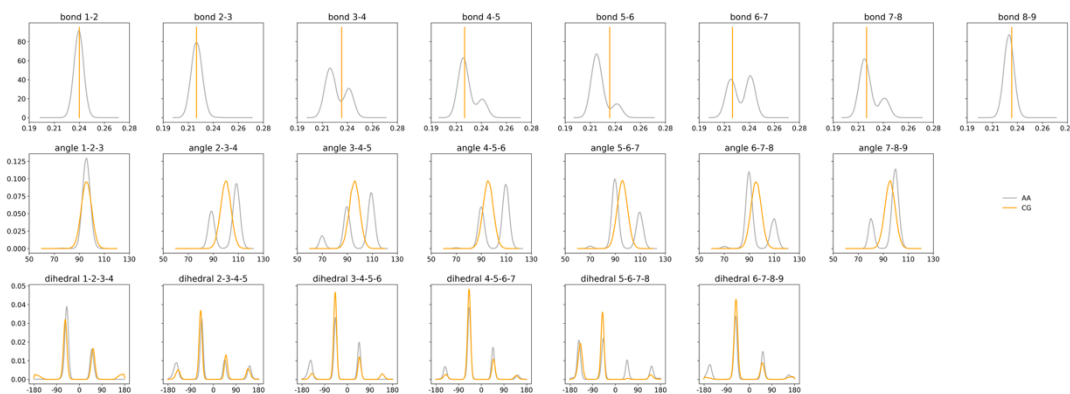


Figure 50: The bonded distribution of PFNA (perfluorononanoic acid).

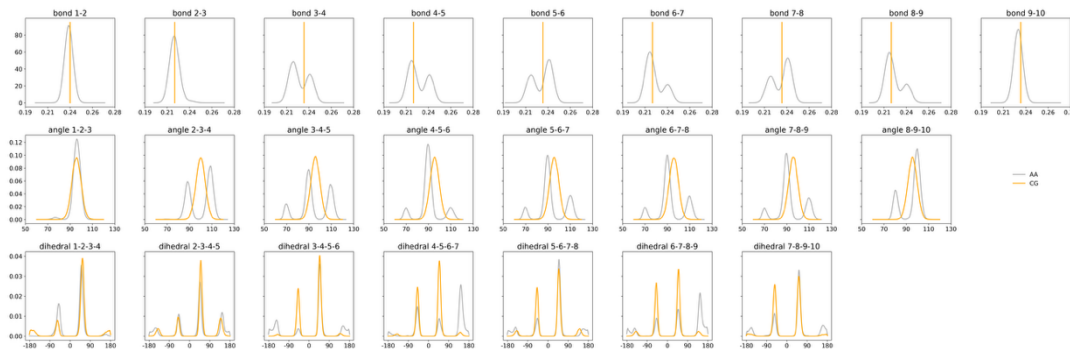


Figure 51: The bonded distribution of PFDA (perfluorodecanoic acid).

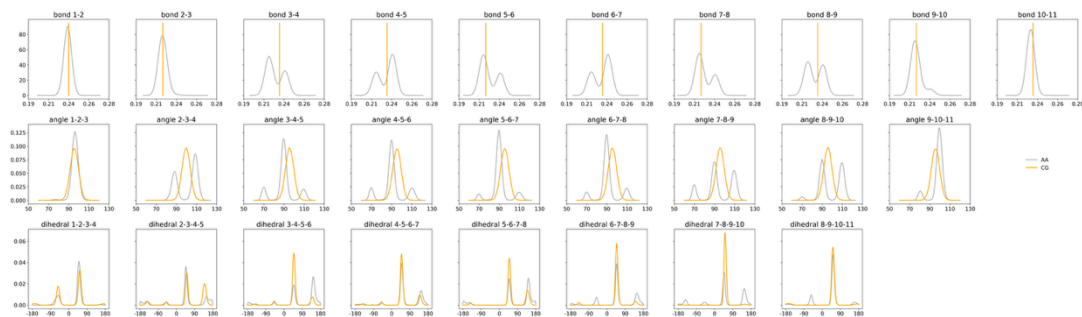


Figure 52: The bonded distribution of PFunDA (perfluoroundecanoic acid).

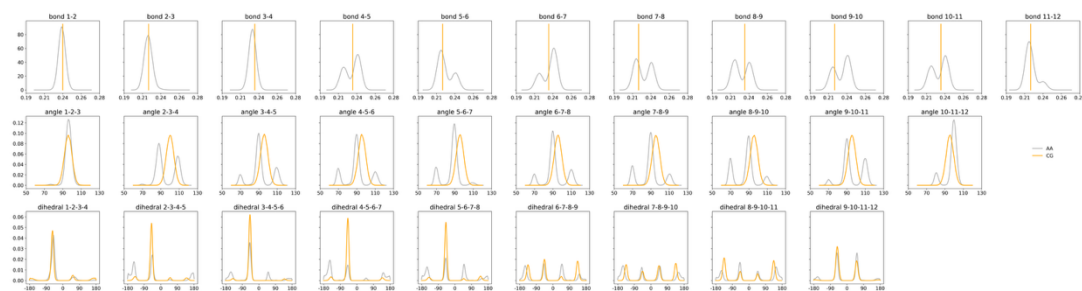


Figure 53: The bonded distribution of PFDODA (perfluorododecanoic acid).

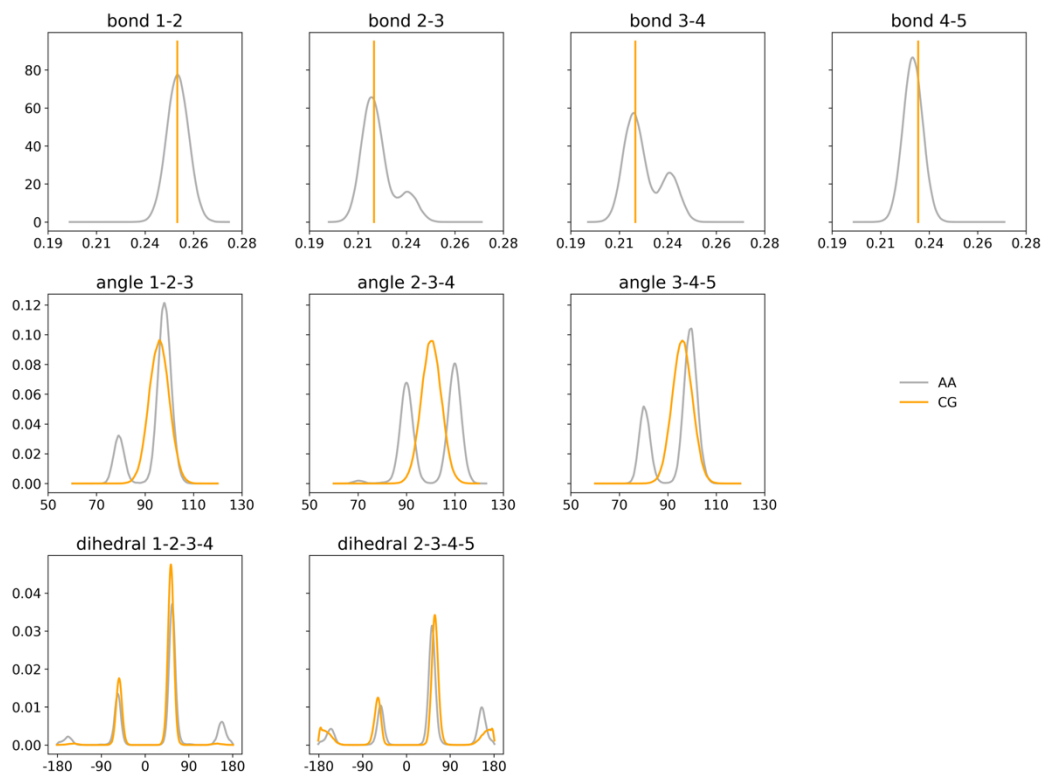


Figure 54: The bonded distribution of PFBS (perfluorobutanesulfonic acid).

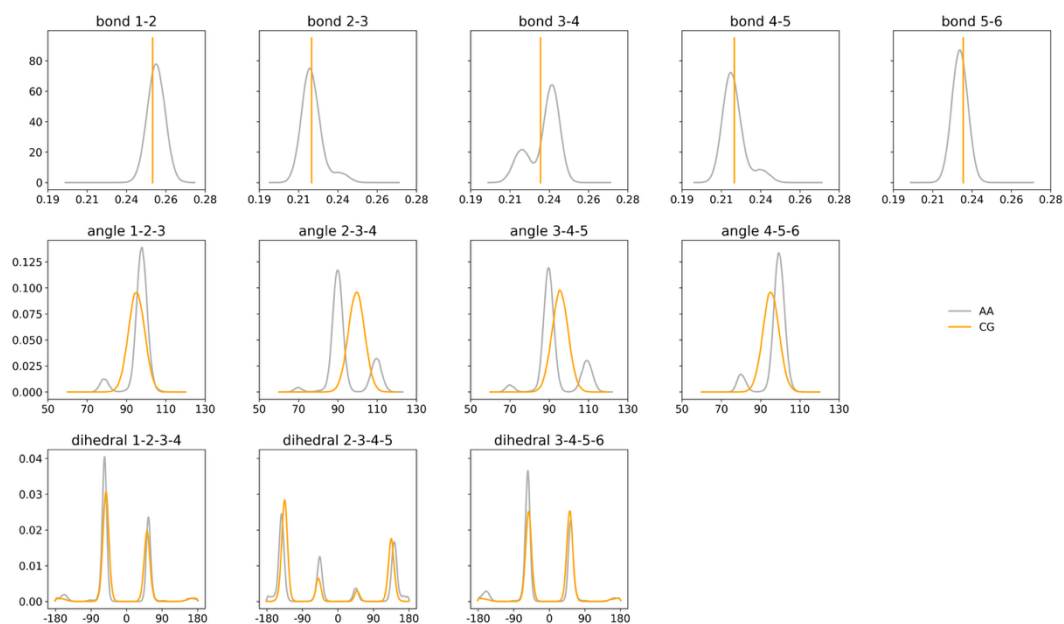


Figure 55: The bonded distribution of PFPeS (perfluoropentanesulfonic acid).

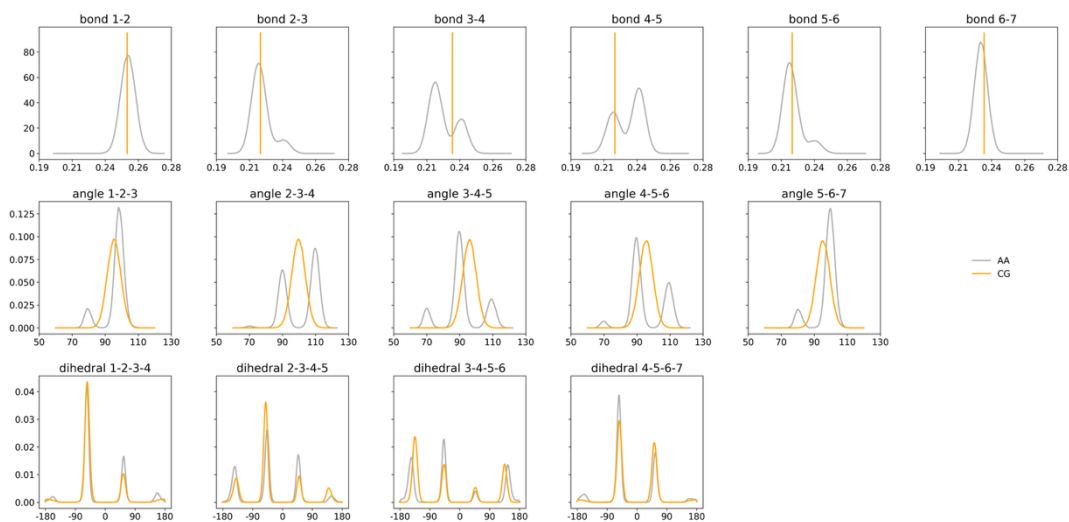


Figure 56: The bonded distribution of PFHxS (perfluorohexanesulfonic acid).

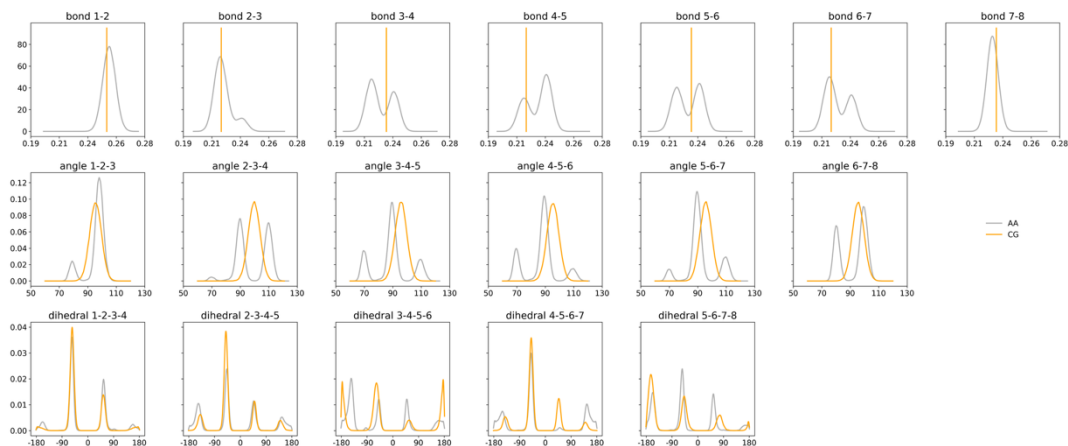


Figure 57: The bonded distribution of PFHpS (perfluoroheptanesulfonic acid).

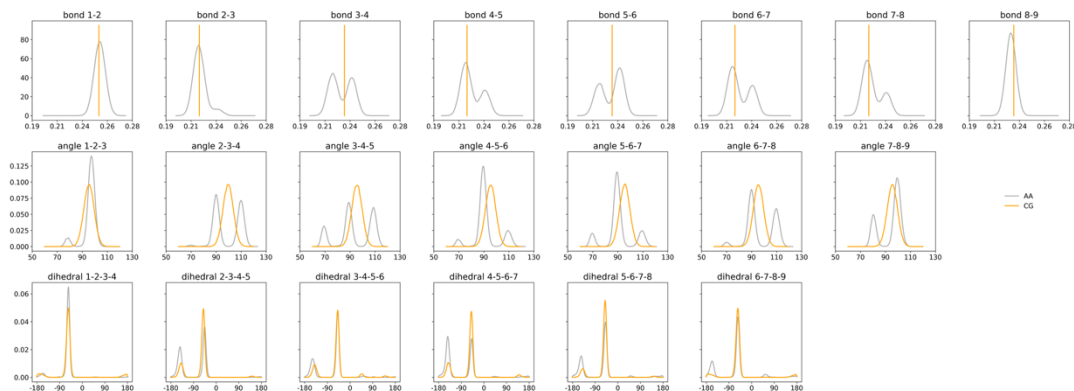


Figure 58: The bonded distribution of PFOS (perfluorooctanesulfonic acid).

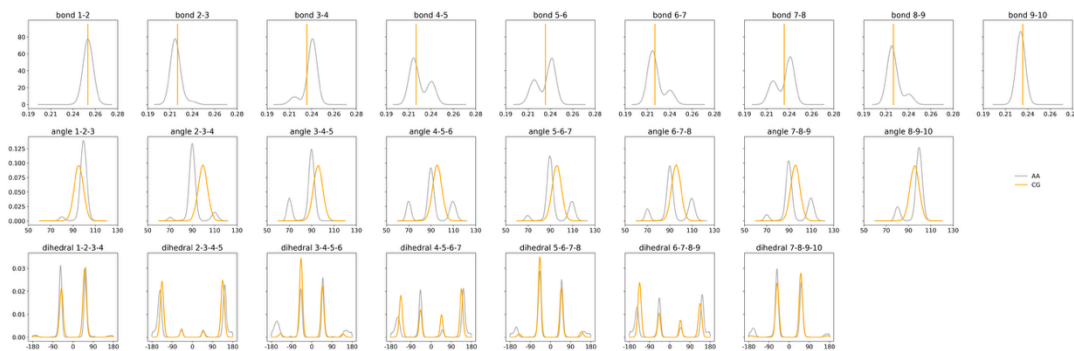


Figure 59: The bonded distribution of PFNS (perfluorononanesulfonic acid).

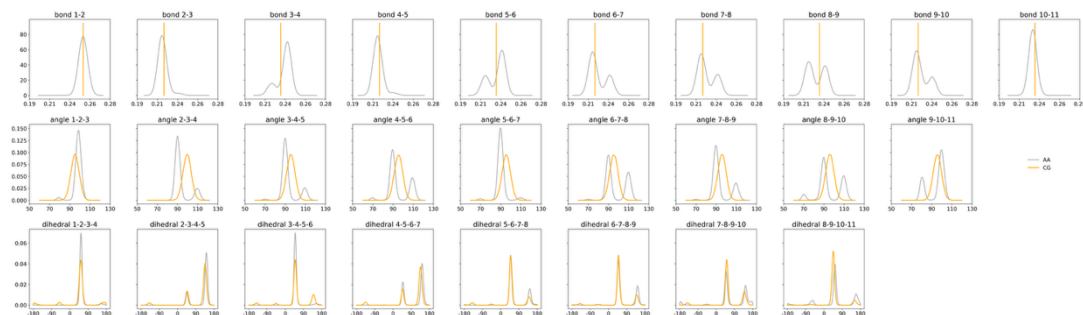


Figure 60: The bonded distribution of PFDS (perfluorodecanesulfonic acid).

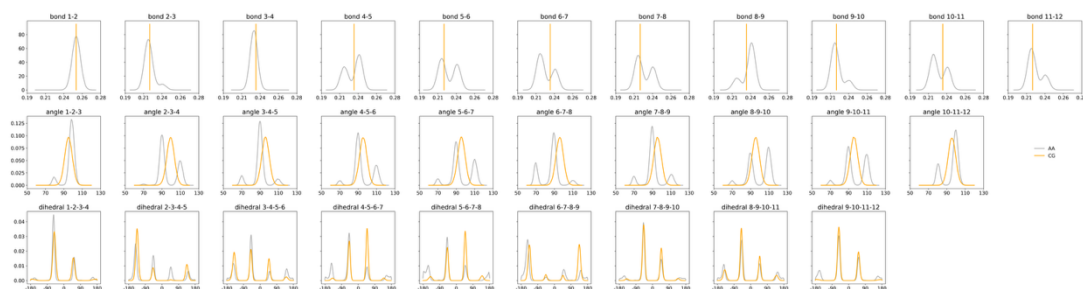


Figure 61: The bonded distribution of PFUnDS (perfluoroundecanesulfonic acid).

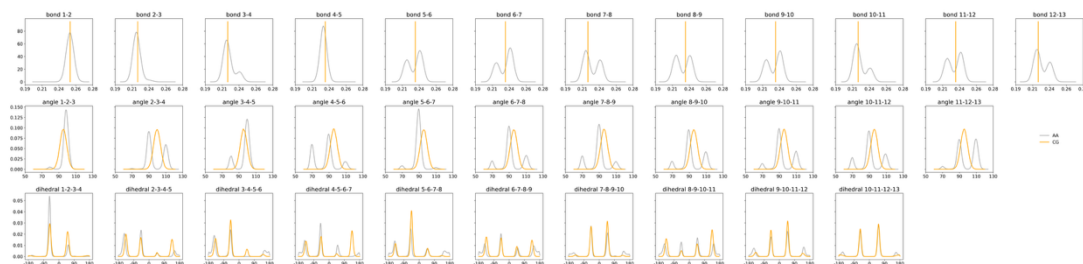


Figure 62: The bonded distribution of PFDoDS (perfluorododecanesulfonic acid).

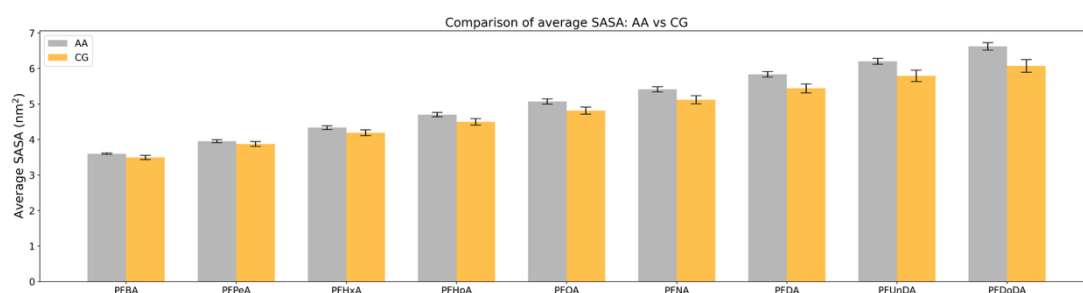


Figure 63: The average SASA distribution of PFCAs.

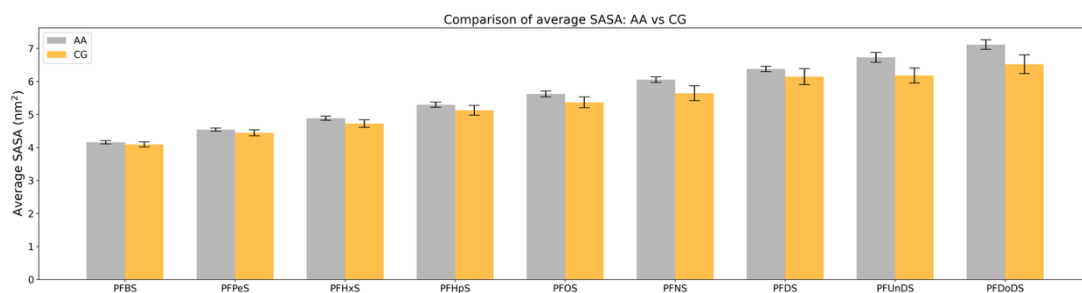


Figure 64: The average SASA distribution of PFASs.

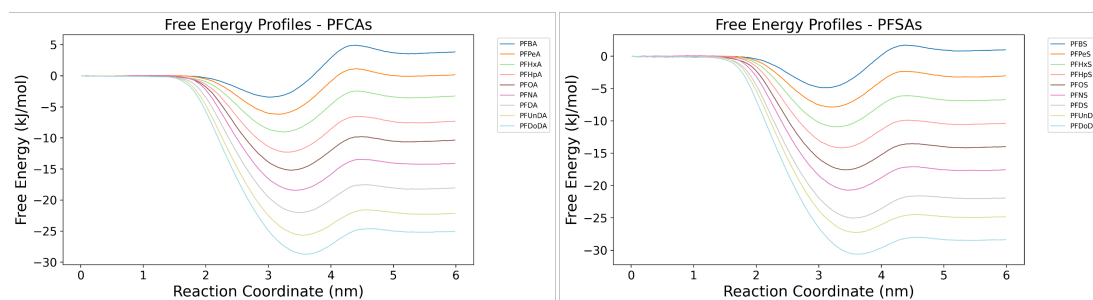


Figure 65: PFM curves of PFCAs and PFASs. The reaction coordinate is defined as the distance between the center of mass (COM) of the water phase and the PFAS molecule. The COM of the octanol phase is at RC=6 nm.

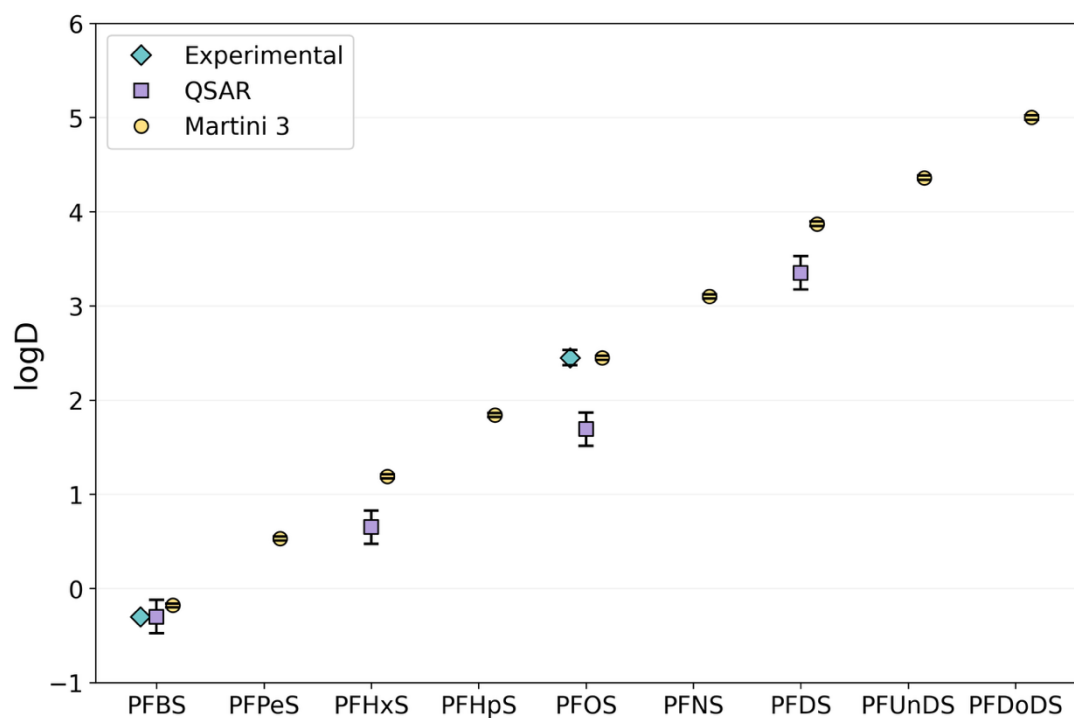


Figure 66: Comparison of PFASs logD values. Values are represented as follows: Martini 3 (circle), experimental (diamond) and QSAR (square) predicted data. When available, error data is reported.

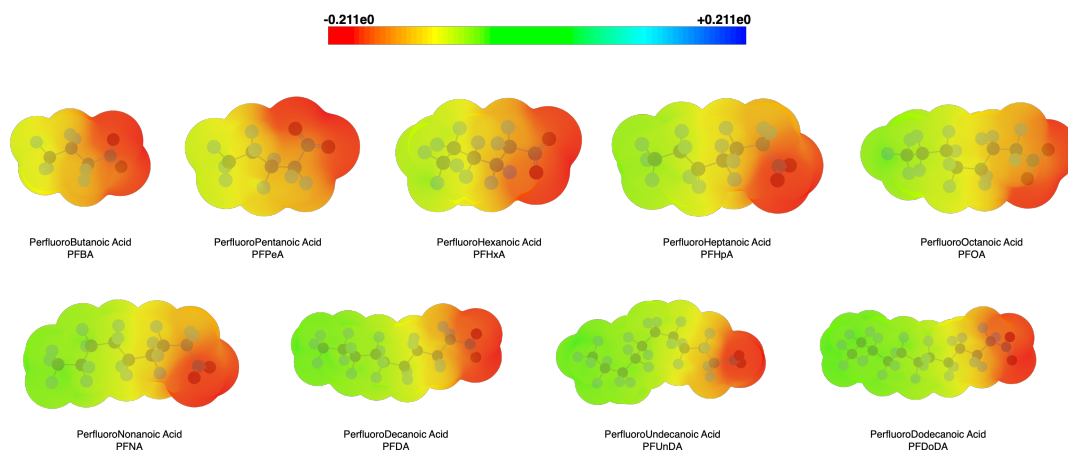


Figure 67: Molecular electrostatic potential maps (MEP) of PFCAs compounds. Red indicates a higher charge density, whereas green indicates a lower charge density.

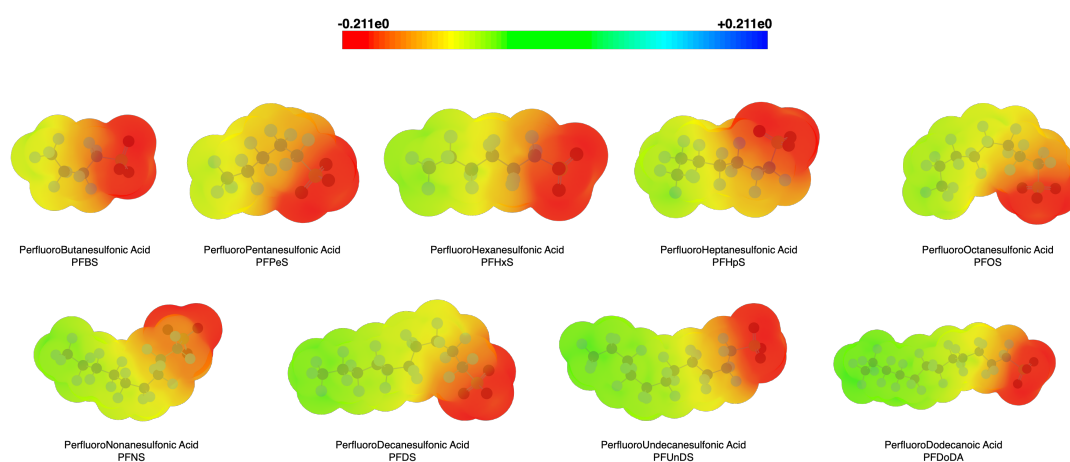


Figure 68: Molecular electrostatic potential maps (MEP) of PFSAs compounds. Red indicates a higher charge density, whereas green indicates a lower charge density.

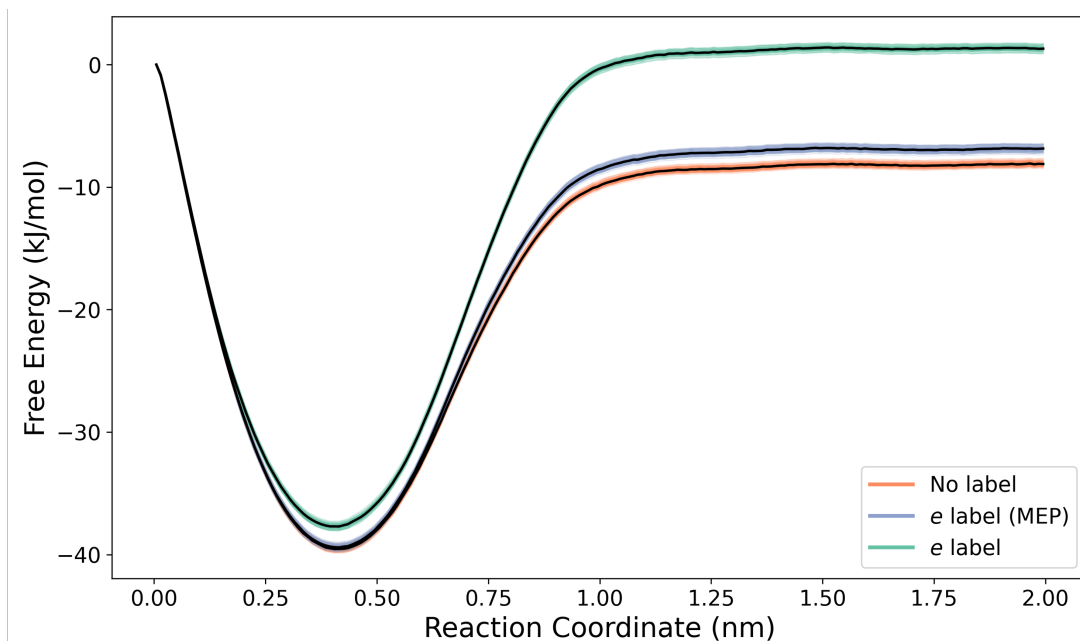


Figure 69: Comparison of PMF curves following the addition of the e electron-rich label to PFOA beads. Labeling was done based on the molecular electrostatic potential map (blue) and considering all the fluorinated chain as electron-rich (green).

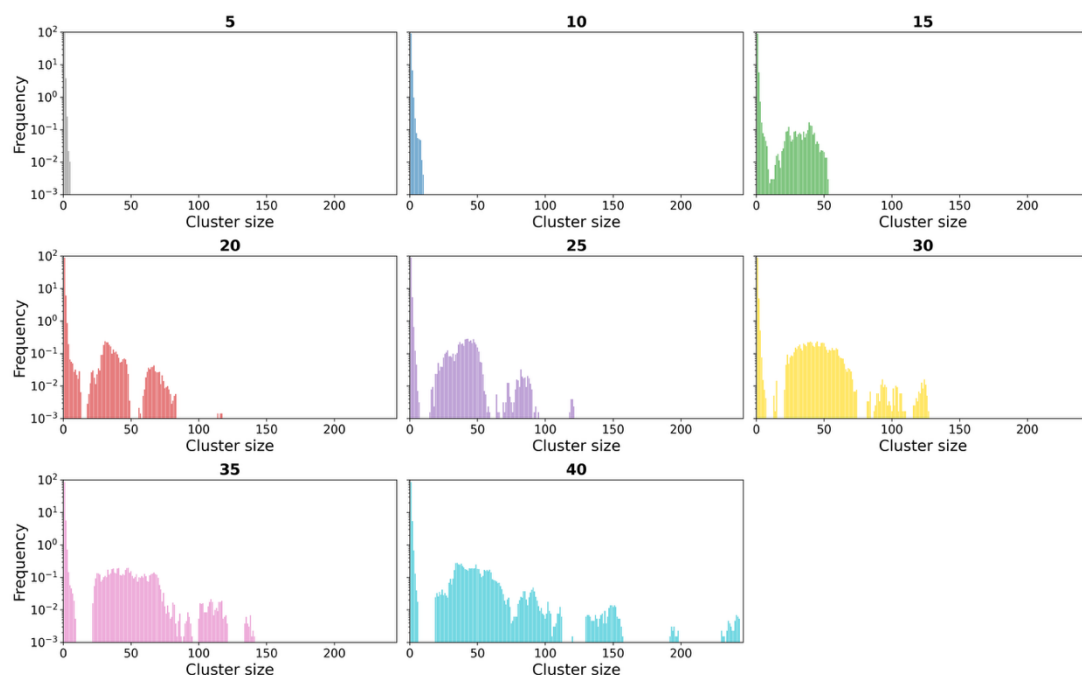


Figure 70: Cluster size distribution for each studied concentration (X beads not labeled) over the last 1 μ s.

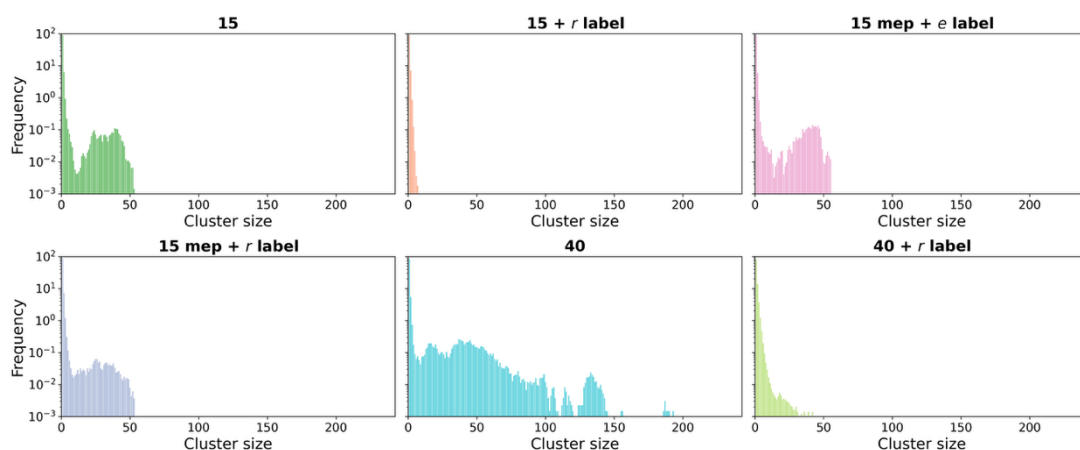


Figure 71: Cluster size distribution for each studied concentration (X beads + r/e label) over the last 1 μ s.

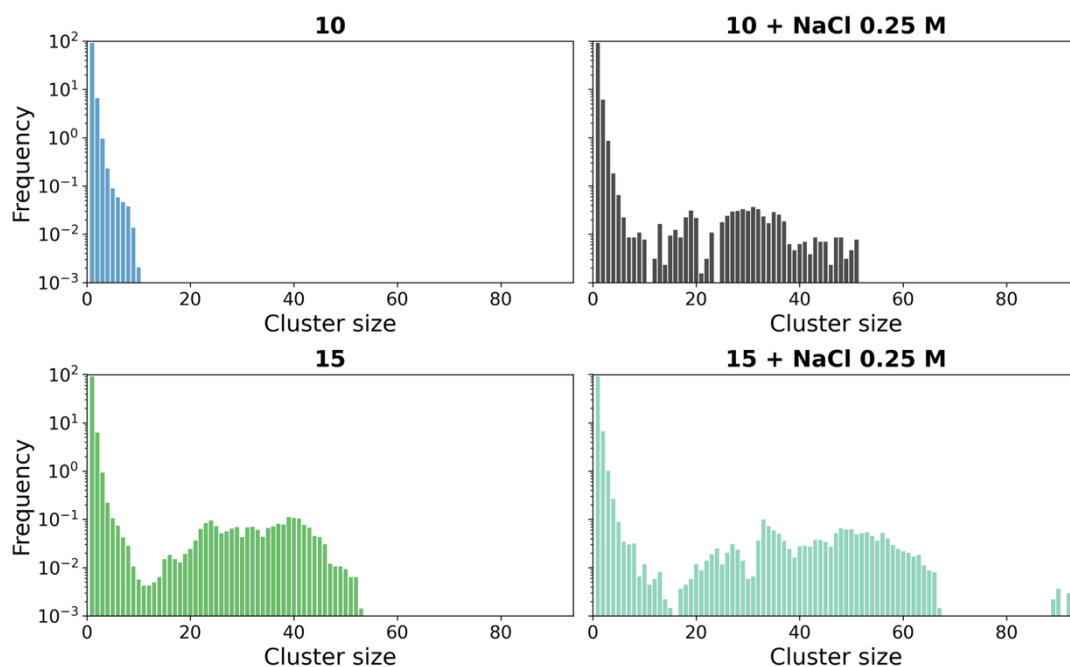


Figure 72: Cluster size distribution for each studied concentration (+ NaCl 0.25 M) over the last 1 μ s.

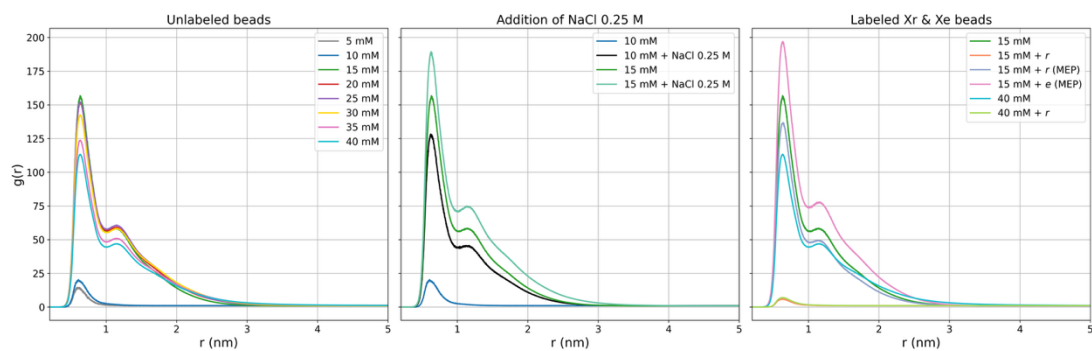


Figure 73: Comparison of the RDF, between the center of mass of each PFOA molecule, for every studied system over the last 1 μ s.

CHAPTER 5: CONCLUSIONS

The present doctoral dissertation examined the extent to which computational protocols originally developed and routinely employed in biophysics can be adapted to address problems in materials science, with a particular focus on wastewater remediation processes. Initially, an *in-house* structural database of emerging pollutants was established on the basis of recent literature by retrieving ASCII molecular descriptors from PubChem. These string-based representations were subsequently translated into three-dimensional molecular structures with explicit consideration of correct stereochemistry and protonation states, thereby yielding a structural emerging contaminants database (EC-DB) comprising more than 2000 compounds, suitable for both biological and environmental applications.

Subsequently, a scalable graphene oxide model was developed by parametrizing individual GO functional groups and assembling them into full-size sheets. A library of four commercially available GO materials was generated, representing commercial samples with distinct oxidation levels and vacancy defect densities. Priority contaminants selected from the EC-DB were then subjected to a virtual screening workflow to predict their adsorption affinities and the underlying driving forces on GOX8-P935 and REGOX8-P980. A grid-based molecular docking protocol, followed by refinement via all-atom molecular dynamics simulations, was employed to exhaustively sample the entire GO surface and elucidate the adsorption mechanisms of emerging contaminants.

The final ranking, obtained using the MM-GBSA method combined with Normal Mode analysis, indicates that the adsorption landscape on graphene oxide is predominantly governed by non-specific dispersive interactions and hydrophobic contacts, which phenomenologically reproduce planar π - π stacking orientations, rather than by electrostatic attraction. This conclusion is corroborated by the higher adsorption affinities calculated for planar polycyclic aromatic hydrocarbons and high molecular weight plasticizers (e.g., DEHP), relative to polar or charged species, which are frequently disfavored due to substantial de-solvation penalties.

Importantly, this study demonstrates that the topological accommodation of the ligand on the heterogeneous GO surface exerts a considerably stronger

influence on binding stability than the absolute stereochemical configuration of the molecule. As observed for chiral contaminants such as ofloxacin and heptachlor, the convergence of binding free energies across distinct stereoisomers indicates that the docking protocol reliably identified stable anchoring modes irrespective of chirality, primarily driven by the capacity of the ligands to conformationally adapt to the surface irregularities.

From a methodological standpoint, the explicit incorporation of entropic contributions proved essential for correcting the intrinsic size-dependence of enthalpy-only scoring functions. The Ligand Efficiency analysis revealed that bulky and highly flexible molecules, such as atorvastatin and DEHP, suffer from considerable thermodynamic penalties arising from the loss of conformational degrees of freedom upon adsorption (rigidification), a physical effect that MM-GBSA alone fails to capture accurately. Furthermore, benchmarking of solvation models identified IGB = 5 (OBC2) as the optimal method for ensuring transferability to future bio-interface applications, while simultaneously highlighting potential artifacts in newer GBn models when applied to perfluorinated compounds. Finally, optimization analyses further substantiated the Quasi-Harmonic approximation as a practical, high-throughput alternative to NMODE for entropy estimation, strictly on the condition that this extensive configurational sampling (>10000 frames) is maintained to achieve robust convergence.

One of the primary aims of this thesis was the rationalization of these molecular interactions to provide practical guidelines for designing future adsorbents. The computational findings clearly indicate that to efficiently remove neutral, planar, and highly hydrophobic pollutants (such as PAHs), filtration materials should maximize the presence of unoxidized, planar graphitic domains, as effectively achieved in reduced GO (rGO). Conversely, for highly polar or anionic species like PFAS, which suffer from pronounced electrostatic repulsion on standard heavily oxidized GO, conventional carbonaceous frameworks are insufficient. In these cases, the predictive models suggest that next-generation adsorbents must incorporate targeted surface functionalization, such as cationic moieties or specific amino acid-derived groups, to provide the necessary electrostatic anchoring points and overcome the energetic barrier of de-solvation.

Crucially, the fragment-based parametrization strategy applied in this work offers a distinct advantage for this optimization phase. By allowing a high-

resolution decomposition of the binding affinity, it enables the precise isolation of energetic contributions from individual surface moieties. While already insightful for characterizing standard GO, this bottom-up thermodynamic mapping will become exceptionally powerful when modeling complex functionalizations like amino acids, as it allows researchers to pinpoint exactly which specific chemical groups drive the adsorption, thereby guiding the rational, atomic-level design of highly selective filtration materials.

In a complementary effort, a linear PFAS library was parametrized using the Martini 3 coarse-grained force field. The resulting coarse-grained model exhibited strong agreement with both reference all-atom simulations and available experimental measurements. In particular, phase behavior and adsorption characteristics on graphene were quantitatively and qualitatively reproduced, thereby underscoring the transferability and robustness of the coarse-grained representation. This parametrized database was subsequently employed to simulate two multicomponent pollutant mixtures designed to approximate realistic environmental scenarios, providing both a conceptual and methodological framework for predictive studies of contaminant transport and competitive adsorption phenomena that would be prohibitively expensive to simulate using conventional AA-MD alone.

Most importantly, the multiscale computational pipeline developed in this work is not restricted to wastewater treatment applications. Rather, it can be generalized to any research context involving small-molecule binding on extended material interfaces, thereby enhancing the accessibility, portability, and broader applicability of virtual screening workflows.

5.1 Limitations and Future Perspectives

While this thesis successfully establishes a robust multiscale *in silico* framework, several methodological and conceptual limitations must be acknowledged to contextualize the findings and guide future research.

First and foremost, while the present predictive framework relies primarily on the internal consistency of its computational protocols, its robustness is already supported by a strong qualitative agreement with experimentally established trends reported in the literature. As highlighted in the results, the model successfully captures the relative sorption affinities of structural analogues (e.g., fluoroquinolones) and accurately predicts the weak interactions of highly

polar or anionic species (such as PFAS and bisphenols), which require specifically functionalized adsorbents. Nevertheless, the current computational results should be interpreted as a robust high-throughput filtering criterion rather than exact thermodynamic representations. Consequently, a primary future objective will be the systematic cross-validation of these theoretical predictions through targeted laboratory-scale adsorption experiments. Building upon the existing literature consensus, this physical validation will be essential to confirm the computational affinities and ensure the models accurately translate to physical filtration systems.

Second, the parametrization of the extended GO surface was conducted using the GAFF force field. While GAFF is highly effective for small organic molecules, it was not originally optimized for macroscopic, heterogeneous 2D materials. Consequently, future studies will aim to test alternative GO parametrizations tailored for nanomaterials. Furthermore, to overcome the limitations of classical force fields in rigorously describing specific electronic effects (such as true π - π stacking), future validation protocols should incorporate QM calculations of molecule-graphene binding energies to benchmark and refine the classical mechanical descriptions.

Regarding the conformational sampling, the current protocol deliberately employed relatively short 10 ns atomistic MD simulations for each of the 18 selected docking poses. This restricted timescale was a methodological choice intended to strictly preserve the spatial distribution generated by the grid-based docking. Extended conventional MD simulations would have inevitably led to ligand diffusion across the graphene basal plane, causing initially distinct poses to drift and converge into the same deep energy minima. By limiting the simulation time to 10 ns, the protocol allowed for the local structural relaxation of the complexes while ensuring a forced, comprehensive sampling of the entire chemically heterogeneous GO surface, including sterically hindered and less energetically favorable regions, thereby preventing pose redundancy.

However, while this strategy was highly effective for evaluating the macroscopic interaction propensity across the whole material, it inherently precluded an exhaustive exploration of the highly complex configurational space and dynamic wrinkling of the flexible GO sheet. Moreover, calculating the reported binding free energies as an unweighted mean across these 18 independent trajectories implicitly assumed that all docking poses, including the less favorable ones, were equally accessible. To balance the need for spatial

sampling with thermodynamic rigor, future simulations must be significantly extended and coupled with enhanced sampling techniques (e.g., Metadynamics). This will allow the system to map the free energy surface without losing spatial control and should be accompanied by the implementation of appropriate statistical weighting strategies (such as Boltzmann weighting) to more accurately reflect the true thermodynamic probability of each distinct binding mode.

Furthermore, it is important to clarify that the initial docking phase served primarily as a high-throughput screening step to eliminate placement bias and identify sensible starting configurations. The core predictive value of this framework lies in the thermodynamic trends extracted after molecular dynamics, and it is these post-MD free energies that ultimately require systematic experimental validation.

Beyond conformational dynamics, the comparative analysis between GOX and REGOX underscored a gap in standard nanosheet modeling. The counterintuitive thermodynamic preference of positively charged compounds for the less electronegative REGOX surface revealed that relying on a single, randomly generated Lerf-Klinowski model introduces significant electrostatic and geometric bias. To better approximate macroscopic experimental reality – where adsorption occurs simultaneously across thousands of inherently heterogeneous nanosheets, future protocols must sample across an ensemble of independent surface replicas – each featuring a unique spatial distribution of defects.

Nonetheless, targeted experimental validation of these post-MD trends will be crucial to properly address the notable dependence of the predicted free energies on the chosen Generalized Born (GB) model. Computationally, this sensitivity can be further mitigated through a consensus scoring strategy: cross-referencing results from multiple docking algorithms, different simulation times, and diverse implicit solvation models will make the pipeline more resilient to force field biases and effectively reduce false positives.

Finally, moving from qualitative observations to predictive design requires a systematic expansion of the computational pipeline. Although the implementation of topological clustering successfully established class-specific adsorption rules, the current dataset of 53 priority contaminants inevitably resulted in a high proportion of structurally unique molecules categorized as

singletons. Populating the chemical space with a broader array of structural analogs is necessary to allow these isolated data points to coalesce into statistically robust clusters, refining the current Structure-Activity Relationship (SAR) insights into comprehensive Quantitative Structure-Property Relationship (QSPR) models.

To support this data scale-up, the construction of the small-molecule database will be fully automated via SMILES enumeration protocols. Simultaneously, the material models will be enriched with complex functional motifs, such as polyethylene glycol (PEG) chains or amino acid derivatives, exploiting the flexibility of the fragment-based parametrization. Similarly, the Martini 3 parameter set will be systematically enlarged to include additional PFAS species and other environmentally relevant contaminants.

By addressing these current limitations and integrating automated high-throughput screening with QSPR analytics, the proposed computational platform will evolve into a highly reliable and systematic tool to guide the rational design and targeted functionalization of next-generation graphene-based nanomaterials.

REFERENCES

- [1] Mekonnen MM, Hoekstra AY. Four billion people facing severe water scarcity. *Science Advances* 2016;2:e1500323. <https://doi.org/10.1126/sciadv.1500323>.
- [2] Oki T, Kanae S. Global Hydrological Cycles and World Water Resources. *Science* 2006;313:1068–72. <https://doi.org/10.1126/science.1128845>.
- [3] Famiglietti JS. The global groundwater crisis. *Nature Clim Change* 2014;4:945–8. <https://doi.org/10.1038/nclimate2425>.
- [4] Trenberth KE, Dai A, van der Schrier G, et al. Global warming and changes in drought. *Nature Clim Change* 2014;4:17–22. <https://doi.org/10.1038/nclimate2067>.
- [5] The Challenge of Micropollutants in Aquatic Systems. *Science* n.d.
- [6] Voulvoulis N, Arpon KD, Giakoumis T. The EU Water Framework Directive: From great expectations to problems with implementation. *Science of The Total Environment* 2017;575:358–66. <https://doi.org/10.1016/j.scitotenv.2016.09.228>.
- [7] COMMISSION IMPLEMENTING DECISION (EU) 2018/ 840 - of 5 June 2018 - establishing a watch list of substances for Union-wide monitoring in the field of water policy pursuant to Directive 2008/105/ EC of the European Parliament and of the Council and repealing Commission Implementing Decision (EU) 2015/ 495 - (notified under document C(2018) 3362) n.d.
- [8] US EPA O. Contaminants of Emerging Concern including Pharmaceuticals and Personal Care Products 2015. <https://www.epa.gov/wqc/contaminants-emerging-concern-including-pharmaceuticals-and-personal-care-products> (accessed February 8, 2026).
- [9] Sauvé S, Desrosiers M. A review of what is an emerging contaminant. *Chemistry Central Journal* 2014;8:15. <https://doi.org/10.1186/1752-153X-8-15>.
- [10] Daughton CG, Ternes TA. Pharmaceuticals and personal care products in the environment: agents of subtle change? *Environ Health Perspect* 1999;107:907–38. <https://doi.org/10.1289/ehp.99107s6907>.
- [11] González-Rubio S, Ballesteros-Gómez A, Asimakopoulos AG, et al. A review on contaminants of emerging concern in European raptors (2002–2020). *Science of The Total Environment* 2021;760:143337. <https://doi.org/10.1016/j.scitotenv.2020.143337>.

- [12] US EPA O. Drinking Water Contaminant Candidate List (CCL) and Regulatory Determination 2014. <https://www.epa.gov/ccl> (accessed February 16, 2026).
- [13] Richardson SD, Ternes TA. Water Analysis: Emerging Contaminants and Current Issues. *Anal Chem* 2022;94:382–416. <https://doi.org/10.1021/acs.analchem.1c04640>.
- [14] Gavrilesco M, Demnerová K, Aamand J, et al. Emerging pollutants in the environment: present and future challenges in biomonitoring, ecological risks and bioremediation. *New Biotechnology* 2015;32:147–56. <https://doi.org/10.1016/j.nbt.2014.01.001>.
- [15] Luo Y, Guo W, Ngo HH, et al. A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment. *Science of The Total Environment* 2014;473–474:619–41. <https://doi.org/10.1016/j.scitotenv.2013.12.065>.
- [16] Kidd KA, Blanchfield PJ, Mills KH, et al. Collapse of a fish population after exposure to a synthetic estrogen. *Proceedings of the National Academy of Sciences* 2007;104:8897–901. <https://doi.org/10.1073/pnas.0609568104>.
- [17] Kumar R, Qureshi M, Vishwakarma DK, et al. A review on emerging water contaminants and the application of sustainable removal technologies. *Case Studies in Chemical and Environmental Engineering* 2022;6:100219. <https://doi.org/10.1016/j.csee.2022.100219>.
- [18] Brodin T, Fick J, Jonsson M, et al. Dilute Concentrations of a Psychiatric Drug Alter Behavior of Fish from Natural Populations. *Science* 2013;339:814–5. <https://doi.org/10.1126/science.1226850>.
- [19] Backhaus T, Faust M. Predictive Environmental Risk Assessment of Chemical Mixtures: A Conceptual Framework. *Environ Sci Technol* 2012;46:2564–73. <https://doi.org/10.1021/es2034125>.
- [20] Heberer T. Occurrence, fate, and removal of pharmaceutical residues in the aquatic environment: a review of recent research data. *Toxicology Letters* 2002;131:5–17. [https://doi.org/10.1016/S0378-4274\(02\)00041-3](https://doi.org/10.1016/S0378-4274(02)00041-3).
- [21] Joss A, Keller E, Alder AC, et al. Removal of pharmaceuticals and fragrances in biological wastewater treatment. *Water Research* 2005;39:3139–52. <https://doi.org/10.1016/j.watres.2005.05.031>.
- [22] Andersson DI, Hughes D. Microbiological effects of sublethal levels of antibiotics. *Nat Rev Microbiol* 2014;12:465–78. <https://doi.org/10.1038/nrmicro3270>.
- [23] Kümmerer K. Antibiotics in the aquatic environment – A review – Part I n.d.

- [24] Ben Y, Fu C, Hu M, et al. Human health risk assessment of antibiotic resistance associated with antibiotic residues in the environment: A review. *Environmental Research* 2019;169:483–93. <https://doi.org/10.1016/j.envres.2018.11.040>.
- [25] Buck RC, Franklin J, Berger U, et al. Perfluoroalkyl and polyfluoroalkyl substances in the environment: Terminology, classification, and origins. *Integrated Environmental Assessment and Management* 2011;7:513–41. <https://doi.org/10.1002/ieam.258>.
- [26] Giesy JP, Kannan K. Global Distribution of Perfluorooctane Sulfonate in Wildlife. *Environ Sci Technol* 2001;35:1339–42. <https://doi.org/10.1021/es001834k>.
- [27] Rahman MF, Peldszus S, Anderson WB. Behaviour and fate of perfluoroalkyl and polyfluoroalkyl substances (PFASs) in drinking water treatment: A review. *Water Research* 2014;50:318–40. <https://doi.org/10.1016/j.watres.2013.10.045>.
- [28] Fenton SE, Ducatman A, Boobis A, et al. Per- and Polyfluoroalkyl Substance Toxicity and Human Health Review: Current State of Knowledge and Strategies for Informing Future Research. *Environmental Toxicology and Chemistry* 2021;40:606–30. <https://doi.org/10.1002/etc.4890>.
- [29] Rivera-Utrilla J, Sánchez-Polo M, Ferro-García MÁ, et al. Pharmaceuticals as emerging contaminants and their removal from water. A review. *Chemosphere* 2013;93:1268–87. <https://doi.org/10.1016/j.chemosphere.2013.07.059>.
- [30] Sharma S, Bhattacharya A. Drinking water contamination and treatment techniques. *Appl Water Sci* 2017;7:1043–67. <https://doi.org/10.1007/s13201-016-0455-7>.
- [31] Snyder SA, Adham S, Redding AM, et al. Role of membranes and activated carbon in the removal of endocrine disruptors and pharmaceuticals. *Desalination* 2007;202:156–81. <https://doi.org/10.1016/j.desal.2005.12.052>.
- [32] Ahammad NA, Ahmad MA, Hameed BH, et al. A mini review of recent progress in the removal of emerging contaminants from pharmaceutical waste using various adsorbents. *Environ Sci Pollut Res* 2023;30:124459–73. <https://doi.org/10.1007/s11356-022-19829-0>.
- [33] Ribeiro AR, Nunes OC, Pereira MFR, et al. An overview on the advanced oxidation processes applied for the treatment of water pollutants defined in the recently launched Directive 2013/39/EU.

- Environment International 2015;75:33–51.
<https://doi.org/10.1016/j.envint.2014.10.027>.
- [34] Guilloso R, Le Roux J, Brosillon S, et al. Benefits of ozonation before activated carbon adsorption for the removal of organic micropollutants from wastewater effluents. Chemosphere 2020;245:125530.
<https://doi.org/10.1016/j.chemosphere.2019.125530>.
- [35] Shannon MA, Bohn PW, Elimelech M, et al. Science and technology for water purification in the coming decades. Nature 2008;452:301–10.
<https://doi.org/10.1038/nature06599>.
- [36] Roduner E. Size matters: why nanomaterials are different. Chem Soc Rev 2006;35:583–92. <https://doi.org/10.1039/B502142C>.
- [37] Qu X, Alvarez PJJ, Li Q. Applications of nanotechnology in water and wastewater treatment. Water Research 2013;47:3931–46.
<https://doi.org/10.1016/j.watres.2012.09.058>.
- [38] Ali I. New Generation Adsorbents for Water Treatment. Chem Rev 2012;112:5073–91. <https://doi.org/10.1021/cr300133d>.
- [39] Mauter MS, Elimelech M. Environmental Applications of Carbon-Based Nanomaterials. Environ Sci Technol 2008;42:5843–59.
<https://doi.org/10.1021/es8006904>.
- [40] Apul OG, Karanfil T. Adsorption of synthetic organic contaminants by carbon nanotubes: A critical review. Water Research 2015;68:34–55.
<https://doi.org/10.1016/j.watres.2014.09.032>.
- [41] Pan B, Xing B. Adsorption Mechanisms of Organic Chemicals on Carbon Nanotubes. Environ Sci Technol 2008;42:9005–13.
<https://doi.org/10.1021/es801777n>.
- [42] Moro G, Khaliha S, Pintus A, et al. Amino acid modified graphene oxide for the simultaneous capture and electrochemical detection of glyphosate. Materials Today Chemistry 2024;36:101936.
<https://doi.org/10.1016/j.mtchem.2024.101936>.
- [43] Geim AK, Novoselov KS. The rise of graphene. Nature Mater 2007;6:183–91. <https://doi.org/10.1038/nmat1849>.
- [44] Lee C, Wei X, Kysar JW, et al. Measurement of the Elastic Properties and Intrinsic Strength of Monolayer Graphene. Science 2008;321:385–8.
<https://doi.org/10.1126/science.1157996>.
- [45] Balandin AA, Ghosh S, Bao W, et al. Superior Thermal Conductivity of Single-Layer Graphene. Nano Lett 2008;8:902–7.
<https://doi.org/10.1021/nl0731872>.

- [46] Nair RR, Blake P, Grigorenko AN, et al. Fine Structure Constant Defines Visual Transparency of Graphene. *Science* 2008;320:1308–1308. <https://doi.org/10.1126/science.1156965>.
- [47] Castro Neto AH, Guinea F, Peres NMR, et al. The electronic properties of graphene. *Rev Mod Phys* 2009;81:109–62. <https://doi.org/10.1103/RevModPhys.81.109>.
- [48] Ferrari AC, Bonaccorso F, Fal’ko V, et al. Science and technology roadmap for graphene, related two-dimensional crystals, and hybrid systems. *Nanoscale* 2015;7:4598–810. <https://doi.org/10.1039/C4NR01600A>.
- [49] Novoselov KS, Fal’ko VI, Colombo L, et al. A roadmap for graphene. *Nature* 2012;490:192–200. <https://doi.org/10.1038/nature11458>.
- [50] Magro MR, Vella DA, Cassar G. Synthesis of graphene oxide: A refined approach. *Carbon Trends* 2025;20:100509. <https://doi.org/10.1016/j.cartre.2025.100509>.
- [51] Dreyer DR, Park S, Bielawski CW, et al. The chemistry of graphene oxide. *Chem Soc Rev* 2009;39:228–40. <https://doi.org/10.1039/B917103G>.
- [52] Jiříčková A, Jankovský O, Sofer Z, et al. Synthesis and Applications of Graphene Oxide. *Materials* 2022;15:920. <https://doi.org/10.3390/ma15030920>.
- [53] Lerf A, He H, Forster M, et al. Structure of Graphite Oxide Revisited. *J Phys Chem B* 1998;102:4477–82. <https://doi.org/10.1021/jp9731821>.
- [54] Li D, Müller MB, Gilje S, et al. Processable aqueous dispersions of graphene nanosheets. *Nature Nanotech* 2008;3:101–5. <https://doi.org/10.1038/nnano.2007.451>.
- [55] Pei S, Cheng H-M. The reduction of graphene oxide. *Carbon* 2012;50:3210–28. <https://doi.org/10.1016/j.carbon.2011.11.010>.
- [56] Cheng M, Yang R, Zhang L, et al. Restoration of graphene from graphene oxide by defect repair. *Carbon* 2012;50:2581–7. <https://doi.org/10.1016/j.carbon.2012.02.016>.
- [57] De Silva KKH, Huang H-H, Joshi R, et al. Restoration of the graphitic structure by defect repair during the thermal reduction of graphene oxide. *Carbon* 2020;166:74–90. <https://doi.org/10.1016/j.carbon.2020.05.015>.
- [58] Talyzin AV, Mercier G, Klechikov A, et al. Brodie vs Hummers graphite oxides for preparation of multi-layered materials. *Carbon* 2017;115:430–40. <https://doi.org/10.1016/j.carbon.2016.12.097>.

- [59] Feicht P, Biskupek J, Gorelik TE, et al. Brodie's or Hummers' Method: Oxidation Conditions Determine the Structure of Graphene Oxide. *Chemistry A European J* 2019;25:8955–9. <https://doi.org/10.1002/chem.201901499>.
- [60] Chernova EA, Gurianov KE, Brotsman VA, et al. Comparative study of transport properties of membranes based on graphene oxide prepared by Brodie and improved Hummers' methods. *Nanosystems: Phys Chem Math* 2023;14:272–8. <https://doi.org/10.17586/2220-8054-2023-14-2-272-278>.
- [61] Sali S, Mackey HR, Abdala AA. Effect of Graphene Oxide Synthesis Method on Properties and Performance of Polysulfone-Graphene Oxide Mixed Matrix Membranes. *Nanomaterials* 2019;9:769. <https://doi.org/10.3390/nano9050769>.
- [62] You S, Luzan SM, Szabó T, et al. Effect of synthesis method on solvation and exfoliation of graphite oxide. *Carbon* 2013;52:171–80. <https://doi.org/10.1016/j.carbon.2012.09.018>.
- [63] Poh HL, Šaněk F, Ambrosi A, et al. Graphenes prepared by Staudenmaier, Hofmann and Hummers methods with consequent thermal exfoliation exhibit very different electrochemical properties. *Nanoscale* 2012;4:3515–22. <https://doi.org/10.1039/C2NR30490B>.
- [64] Botas C, Álvarez P, Blanco P, et al. Graphene materials with different structures prepared from the same graphite by the Hummers and Brodie methods. *Carbon* 2013;65:156–64. <https://doi.org/10.1016/j.carbon.2013.08.009>.
- [65] Khaliha S, Marforio TD, Kovtun A, et al. Defective graphene nanosheets for drinking water purification: Adsorption mechanism, performance, and recovery. *FlatChem* 2021;29:100283. <https://doi.org/10.1016/j.flatc.2021.100283>.
- [66] Lancellotti L, Bianchi A, Kovtun A, et al. Selective ion transport in large-area graphene oxide membrane filters driven by the ionic radius and electrostatic interactions. *Nanoscale* 2024;16:7123–33. <https://doi.org/10.1039/D3NR05874C>.
- [67] Zhao G, Li J, Ren X, et al. Few-Layered Graphene Oxide Nanosheets As Superior Sorbents for Heavy Metal Ion Pollution Management. *Environ Sci Technol* 2011;45:10454–62. <https://doi.org/10.1021/es203439v>.
- [68] Wang J, Chen B. Adsorption and coadsorption of organic pollutants and a heavy metal by graphene oxide and reduced graphene materials. *Chemical Engineering Journal* 2015;281:379–88. <https://doi.org/10.1016/j.cej.2015.06.102>.

- [69] Gao Y, Li Y, Zhang L, et al. Adsorption and removal of tetracycline antibiotics from aqueous solution by graphene oxide. *Journal of Colloid and Interface Science* 2012;368:540–6. <https://doi.org/10.1016/j.jcis.2011.11.015>.
- [70] Pei Z, Li L, Sun L, et al. Adsorption characteristics of 1,2,4-trichlorobenzene, 2,4,6-trichlorophenol, 2-naphthol and naphthalene on graphene and graphene oxide. *Carbon* 2013;51:156–63. <https://doi.org/10.1016/j.carbon.2012.08.024>.
- [71] Kyzas GZ, Deliyanni EA, Matis KA. Graphene oxide and its application as an adsorbent for wastewater treatment. *Journal of Chemical Technology & Biotechnology* 2014;89:196–205. <https://doi.org/10.1002/jctb.4220>.
- [72] Plastiras O-E, Deliyanni E, Samanidou V. Applications of Graphene-Based Nanomaterials in Environmental Analysis. *Applied Sciences* 2021;11:3028. <https://doi.org/10.3390/app11073028>.
- [73] Boulanger N, Kuzenkova AS, Iakunkov A, et al. Enhanced Sorption of Radionuclides by Defect-Rich Graphene Oxide. *ACS Appl Mater Interfaces* 2020;12:45122–35. <https://doi.org/10.1021/acsami.0c11122>.
- [74] Zakaria ABM, Vasquez ES, Walters KB, et al. Functional holey graphene oxide: a new electrochemically transformed substrate material for dopamine sensing. *RSC Adv* 2015;5:107123–35. <https://doi.org/10.1039/C5RA19991C>.
- [75] Chandra V, Park J, Chun Y, et al. Water-Dispersible Magnetite-Reduced Graphene Oxide Composites for Arsenic Removal. *ACS Nano* 2010;4:3979–86. <https://doi.org/10.1021/nn1008897>.
- [76] Zhang H, Lv X, Li Y, et al. P25-Graphene Composite as a High Performance Photocatalyst. *ACS Nano* 2010;4:380–6. <https://doi.org/10.1021/nn901221k>.
- [77] Pedrosa M, Da Silva ES, Pastrana-Martínez LM, et al. Hummers' and Brodie's graphene oxides as photocatalysts for phenol degradation. *Journal of Colloid and Interface Science* 2020;567:243–55. <https://doi.org/10.1016/j.jcis.2020.01.093>.
- [78] Li Y, Pathak B, Nisar J, et al. Metal-decorated graphene oxide for ammonia adsorption. *EPL* 2013;103:28007. <https://doi.org/10.1209/0295-5075/103/28007>.
- [79] Hu X, Zhou Q. Health and Ecosystem Risks of Graphene. *Chem Rev* 2013;113:3815–35. <https://doi.org/10.1021/cr300045n>.

- [80] Fadeel B, Bussy C, Merino S, et al. Safety Assessment of Graphene-Based Materials: Focus on Human Health and the Environment. *ACS Nano* 2018;12:10582–620. <https://doi.org/10.1021/acsnano.8b04758>.
- [81] Holt BD, Arnold AM, Sydlík SA. In It for the Long Haul: The Cytocompatibility of Aged Graphene Oxide and Its Degradation Products. *Adv Healthcare Materials* 2016;5:3056–66. <https://doi.org/10.1002/adhm.201600745>.
- [82] Potts JR, Dreyer DR, Bielawski CW, et al. Graphene-based polymer nanocomposites. *Polymer* 2011;52:5–25. <https://doi.org/10.1016/j.polymer.2010.11.042>.
- [83] Tunioli F, Khaliha S, Mantovani S, et al. Adsorption of emerging contaminants by graphene related materials and their alginate composite hydrogels. *Journal of Environmental Chemical Engineering* 2023;11:109566. <https://doi.org/10.1016/j.jece.2023.109566>.
- [84] Yin J, Deng B. Polymer-matrix nanocomposite membranes for water treatment. *Journal of Membrane Science* 2015;479:256–75. <https://doi.org/10.1016/j.memsci.2014.11.019>.
- [85] Kovtun A, Bianchi A, Zambianchi M, et al. Core-shell graphene oxide-polymer hollow fibers as water filters with enhanced performance and selectivity. *Faraday Discuss* 2021;227:274–90. <https://doi.org/10.1039/C9FD00117D>.
- [86] Zubair M, Farooq S, Hussain A, et al. A review of current developments in graphene oxide-polysulfone derived membranes for water remediation. *Environ Sci: Adv* 2024;3:983–1003. <https://doi.org/10.1039/D4VA00058G>.
- [87] Khaliha S, Tunioli F, Foti L, et al. Upcycling of plastic membrane industrial scraps and reuse as sorbent for emerging contaminants in water. *Environ Sci: Water Res Technol* 2024;10:1097–107. <https://doi.org/10.1039/D3EW00900A>.
- [88] Wang Y, Xiao J, Suzek TO, et al. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 2009;37:W623–33. <https://doi.org/10.1093/nar/gkp456>.
- [89] Kim S, Chen J, Cheng T, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res* 2021;49:D1388–95. <https://doi.org/10.1093/nar/gkaa971>.
- [90] Kim S, Thiessen PA, Bolton EE, et al. PubChem Substance and Compound databases. *Nucleic Acids Res* 2016;44:D1202–13. <https://doi.org/10.1093/nar/gkv951>.

- [91] Weininger D, Weininger A, Weininger JL. SMILES. 2. Algorithm for generation of unique SMILES notation. *J Chem Inf Comput Sci* 1989;29:97–101. <https://doi.org/10.1021/ci00062a008>.
- [92] Heller SR, McNaught A, Pletnev I, et al. InChI, the IUPAC International Chemical Identifier. *J Cheminform* 2015;7:23. <https://doi.org/10.1186/s13321-015-0068-4>.
- [93] Southan C. InChI in the wild: an assessment of InChIKey searching in Google. *J Cheminform* 2013;5:10. <https://doi.org/10.1186/1758-2946-5-10>.
- [94] Cheng T, Zhao Y, Li X, et al. Computation of Octanol–Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J Chem Inf Model* 2007;47:2140–8. <https://doi.org/10.1021/ci700257y>.
- [95] Born M, Oppenheimer R. Zur Quantentheorie der Molekeln. *Annalen Der Physik* 1927;389:457–84. <https://doi.org/10.1002/andp.19273892002>.
- [96] Roothaan CCJ. New Developments in Molecular Orbital Theory. *Rev Mod Phys* 1951;23:69–89. <https://doi.org/10.1103/RevModPhys.23.69>.
- [97] Hehre WJ, Ditchfield R, Pople JA. Self—Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J Chem Phys* 1972;56:2257–61. <https://doi.org/10.1063/1.1677527>.
- [98] Hariharan PC, Pople JA. The influence of polarization functions on molecular orbital hydrogenation energies. *Theoret Chim Acta* 1973;28:213–22. <https://doi.org/10.1007/BF00533485>.
- [99] Clark T, Chandrasekhar J, Spitznagel GW, et al. Efficient diffuse function-augmented basis sets for anion calculations. III. The 3-21+G basis set for first-row elements, Li–F. *Journal of Computational Chemistry* 1983;4:294–301. <https://doi.org/10.1002/jcc.540040303>.
- [100] Case DA, Aktulga HM, Belfon K, et al. Amber 2025. University of California, San Francisco; 2025.
- [101] Case DA, Cerutti DS, Cruzeiro VWD, et al. Recent Developments in Amber Biomolecular Simulations. *J Chem Inf Model* 2025;65:7835–43. <https://doi.org/10.1021/acs.jcim.5c01063>.
- [102] Case DA, Aktulga HM, Belfon K, et al. AmberTools. *J Chem Inf Model* 2023;63:6183–91. <https://doi.org/10.1021/acs.jcim.3c01153>.
- [103] Wang J, Wolf RM, Caldwell JW, et al. Development and testing of a general amber force field. *Journal of Computational Chemistry* 2004;25:1157–74. <https://doi.org/10.1002/jcc.20035>.

- [104] Wang J, Wang W, Kollman PA, et al. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling* 2006;25:247–60. <https://doi.org/10.1016/j.jmgm.2005.12.005>.
- [105] Bayly CI, Cieplak P, Cornell W, et al. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 1993;97:10269–80. <https://doi.org/10.1021/j100142a004>.
- [106] Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian 16 Rev. C.01 2016.
- [107] Cornell WD, Cieplak P, Bayly CI, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J Am Chem Soc* 1995;117:5179–97. <https://doi.org/10.1021/ja00124a002>.
- [108] Besler BH, Merz Jr. KM, Kollman PA. Atomic charges derived from semiempirical methods. *Journal of Computational Chemistry* 1990;11:431–9. <https://doi.org/10.1002/jcc.540110404>.
- [109] Yang Y, Zhang X, Jiang J, et al. Which Micropollutants in Water Environments Deserve More Attention Globally? *Environ Sci Technol* 2022;56:13–29. <https://doi.org/10.1021/acs.est.1c04250>.
- [110] Landrum G, Tosco P, Kelley B, et al. rdkit/rdkit: 2025_09_6 (Q3 2025) Release (Release_2025_09_6) 2026. <https://doi.org/10.5281/zenodo.18797641>.
- [111] ChemAxon Ltd. pKa calculator plugin 2026.
- [112] Schwarzenbach RP, Gschwend PM, Imboden DM, et al. *Environmental Organic Chemistry*, 2nd Edition n.d.
- [113] Bagnis S, Fitzsimons MF, Snape J, et al. Processes of distribution of pharmaceuticals in surface freshwaters: implications for risk assessment. *Environ Chem Lett* 2018;16:1193–216. <https://doi.org/10.1007/s10311-018-0742-7>.
- [114] Da Rocha Medeiros G, Da Silva Pereira Júnior A, Galvão FMF, et al. Optimization of diclofenac sodium adsorption onto graphene nanosheets: capacity, kinetics, isotherms and removal. *Desalination and Water Treatment* 2022;271:176–91. <https://doi.org/10.5004/dwt.2022.28789>.
- [115] Rocha Medeiros G da, de Lima PR, Júnior C de AM, et al. Diclofenac sodium removal by powdered activated carbon of coconut endocarp: process optimization, kinetics, and isotherms. *Water Practice and Technology* 2024;19:4906–27. <https://doi.org/10.2166/wpt.2024.281>.
- [116] Tebes-Stevens C, Patel JM, Jones WJ, et al. Prediction of Hydrolysis Products of Organic Chemicals under Environmental pH Conditions.

- Environ Sci Technol 2017;51:5008–16. <https://doi.org/10.1021/acs.est.6b05412>.
- [117] Turner GSC, Mills GA, Teasdale PR, et al. Evaluation of DGT techniques for measuring inorganic uranium species in natural waters: Interferences, deployment time and speciation. *Analytica Chimica Acta* 2012;739:37–46. <https://doi.org/10.1016/j.aca.2012.06.011>.
- [118] Dennington R, Keith T, Millam J. GaussView, Version 6.1.1 2019.
- [119] Hanwell MD, Curtis DE, Lonie DC, et al. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J Cheminform* 2012;4:17. <https://doi.org/10.1186/1758-2946-4-17>.
- [120] O’Boyle NM, Banck M, James CA, et al. Open Babel: An open chemical toolbox. *J Cheminform* 2011;3:33. <https://doi.org/10.1186/1758-2946-3-33>.
- [121] Rappe AK, Casewit CJ, Colwell KS, et al. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J Am Chem Soc* 1992;114:10024–35. <https://doi.org/10.1021/ja00051a040>.
- [122] LayerOne AS. Technical data sheet: GOX8® graphene oxide P935. Oslo, Norway: LayerOne AS; 2026.
- [123] LayerOne AS. Technical data sheet: REGOX8® graphene P980. Oslo, Norway: LayerOne AS; 2026.
- [124] LayerOne AS. Technical data sheet: REGOX8® graphene P995. Oslo, Norway: LayerOne AS; 2026.
- [125] LayerOne AS. Technical data sheet: REGOX8® graphene P999. Oslo, Norway: LayerOne AS; 2026.
- [126] Stauffer D, Dragneva N, Floriano WB, et al. An atomic charge model for graphene oxide for exploring its bioadhesive properties in explicit water. *J Chem Phys* 2014;141:044705. <https://doi.org/10.1063/1.4890503>.
- [127] Smith AP, McKercher AE, Mawhinney RC. Inductive Effect: A Quantum Theory of Atoms in Molecules Perspective. *J Phys Chem A* 2011;115:12544–54. <https://doi.org/10.1021/jp202757p>.
- [128] Liu Y, Wilcox J. Effects of Surface Heterogeneity on the Adsorption of CO₂ in Microporous Carbons. *Environ Sci Technol* 2012;46:1940–7. <https://doi.org/10.1021/es204071g>.
- [129] Garcia NA, Awuah JB, Zhao C, et al. Simulation-ready graphene oxide structures with hierarchical complexity: a modular tiling strategy. *2D Mater* 2023;10:025007. <https://doi.org/10.1088/2053-1583/acb0e1>.

- [130] Muraru S, Burns JS, Ionita M. GOPY: A tool for building 2D graphene-based computational models. *SoftwareX* 2020;12:100586. <https://doi.org/10.1016/j.softx.2020.100586>.
- [131] Bagri A, Mattevi C, Acik M, et al. Structural evolution during the reduction of chemically derived graphene oxide. *Nature Chem* 2010;2:581–7. <https://doi.org/10.1038/nchem.686>.
- [132] Erickson K, Erni R, Lee Z, et al. Determination of the Local Chemical Structure of Graphene Oxide and Reduced Graphene Oxide. *Advanced Materials* 2010;22:4467–72. <https://doi.org/10.1002/adma.201000732>.
- [133] Pintus A, Mantovani S, Kovtun A, et al. Recyclable GO-Arginine Hybrids for CO₂ Fixation into Cyclic Carbonates. *Chemistry – A European Journal* 2023;29:e202202440. <https://doi.org/10.1002/chem.202202440>.
- [134] Mantovani S, Marforio TD, Khaliha S, et al. Amino acid-driven adsorption of emerging contaminants in water by modified graphene oxide nanosheets. *Environ Sci: Water Res Technol* 2023;9:1030–40. <https://doi.org/10.1039/D2EW00871H>.
- [135] Konkena B, Vasudevan S. Understanding Aqueous Dispersibility of Graphene Oxide and Reduced Graphene Oxide through pK_a Measurements. *J Phys Chem Lett* 2012;3:867–72. <https://doi.org/10.1021/jz300236w>.
- [136] Dimiev AM, Alemany LB, Tour JM. Graphene Oxide. Origin of Acidity, Its Instability in Water, and a New Dynamic Structural Model. *ACS Nano* 2013;7:576–88. <https://doi.org/10.1021/nn3047378>.
- [137] Oprea TI. Virtual Screening in Lead Discovery: A Viewpoint. *Molecules* 2002;7:51–62. <https://doi.org/10.3390/70100051>.
- [138] Santana K, do Nascimento LD, Lima e Lima A, et al. Applications of Virtual Screening in Bioprospecting: Facts, Shifts, and Perspectives to Explore the Chemo-Structural Diversity of Natural Products. *Front Chem* 2021;9. <https://doi.org/10.3389/fchem.2021.662688>.
- [139] Best Practices for Docking-Based Virtual Screening. *Molecular Docking for Computer-Aided Drug Design*, Academic Press; 2021, p. 75–98. <https://doi.org/10.1016/B978-0-12-822312-3.00001-1>.
- [140] Wang G, Zhu W. Molecular Docking for Drug Discovery and Development: A Widely Used Approach but Far from Perfect. *Future Medicinal Chemistry* 2016;8:1707–10. <https://doi.org/10.4155/fmc-2016-0143>.
- [141] Dias R, Jr WF de A. Molecular Docking Algorithms. *Current Drug Targets* n.d.;9:1040–7. <https://doi.org/10.2174/138945008786949432>.

- [142] Use of molecular docking computational tools in drug discovery. *Progress in Medicinal Chemistry*, vol. 60, Elsevier; 2021, p. 273–343. <https://doi.org/10.1016/bs.pmch.2021.01.004>.
- [143] Torres PHM, Sodero ACR, Jofily P, et al. Key Topics in Molecular Docking for Drug Design. *International Journal of Molecular Sciences* 2019;20:4574. <https://doi.org/10.3390/ijms20184574>.
- [144] Koukos PI, Réau M, Bonvin AMJJ. Shape-Restrained Modeling of Protein–Small-Molecule Complexes with High Ambiguity Driven DOCKing. *J Chem Inf Model* 2021;61:4807–18. <https://doi.org/10.1021/acs.jcim.1c00796>.
- [145] Yan C, Xu X, Zou X. Fully Blind Docking at the Atomic Level for Protein-Peptide Complex Structure Prediction. *Structure* 2016;24:1842–53. <https://doi.org/10.1016/j.str.2016.07.021>.
- [146] Zhou P, Jin B, Li H, et al. HPEPDOCK: a web server for blind peptide–protein docking based on a hierarchical algorithm. *Nucleic Acids Research* 2018;46:W443–50. <https://doi.org/10.1093/nar/gky357>.
- [147] Ugurlu SY, McDonald D, Lei H, et al. Cobdock: an accurate and practical machine learning-based consensus blind docking method. *J Cheminform* 2024;16:5. <https://doi.org/10.1186/s13321-023-00793-x>.
- [148] Meng X-Y, Zhang H-X, Mezei M, et al. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer-Aided Drug Design* n.d.;7:146–57. <https://doi.org/10.2174/157340911795677602>.
- [149] Ferreira LG, Dos Santos RN, Oliva G, et al. Molecular Docking and Structure-Based Drug Design Strategies. *Molecules* 2015;20:13384–421. <https://doi.org/10.3390/molecules200713384>.
- [150] Halperin I, Ma B, Wolfson H, et al. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Structure, Function, and Bioinformatics* 2002;47:409–43. <https://doi.org/10.1002/prot.10115>.
- [151] Lexa KW, Carlson HA. Protein flexibility in docking and surface mapping. *Quarterly Reviews of Biophysics* 2012;45:301–43. <https://doi.org/10.1017/S0033583512000066>.
- [152] Sherman W, Day T, Jacobson MP, et al. Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J Med Chem* 2006;49:534–53. <https://doi.org/10.1021/jm050540c>.
- [153] Morris GM, Lim-Wilby M. Molecular Docking. In: Kukol A, editor. *Molecular Modeling of Proteins*, Totowa, NJ: Humana Press; 2008, p. 365–82. https://doi.org/10.1007/978-1-59745-177-2_19.

- [154] Kitchen DB, Decornez H, Furr JR, et al. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004;3:935–49. <https://doi.org/10.1038/nrd1549>.
- [155] Sousa SF, Fernandes PA, Ramos MJ. Protein–ligand docking: Current status and future challenges. *Proteins: Structure, Function, and Bioinformatics* 2006;65:15–26. <https://doi.org/10.1002/prot.21082>.
- [156] Huang S-Y, Grinter SZ, Zou X. Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions. *Phys Chem Chem Phys* 2010;12:12899. <https://doi.org/10.1039/c0cp00151a>.
- [157] Böhm H-J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Computer-Aided Mol Des* 1994;8:243–56. <https://doi.org/10.1007/BF00126743>.
- [158] Gilson MK, Zhou H-X. Calculation of Protein-Ligand Binding Affinities*. *Annual Review of Biophysics* 2007;36:21–42. <https://doi.org/10.1146/annurev.biophys.36.040306.132550>.
- [159] Liu J, Wang R. Classification of Current Scoring Functions. *J Chem Inf Model* 2015;55:475–82. <https://doi.org/10.1021/ci500731a>.
- [160] Eberhardt J, Santos-Martins D, Tillack AF, et al. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J Chem Inf Model* 2021;61:3891–8. <https://doi.org/10.1021/acs.jcim.1c00203>.
- [161] Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* 2010;31:455–61. <https://doi.org/10.1002/jcc.21334>.
- [162] Morris GM, Huey R, Lindstrom W, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* 2009;30:2785–91. <https://doi.org/10.1002/jcc.21256>.
- [163] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 1996;14:33–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [164] Karplus M, Petsko GA. Molecular dynamics simulations in biology. *Nature* 1990;347:631–9. <https://doi.org/10.1038/347631a0>.
- [165] Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. *Neuron* 2018;99:1129–43. <https://doi.org/10.1016/j.neuron.2018.08.011>.

- [166] Ponder JW, Case DA. Force Fields for Protein Simulations. *Advances in Protein Chemistry*, vol. 66, Elsevier; 2003, p. 27–85. [https://doi.org/10.1016/S0065-3233\(03\)66002-X](https://doi.org/10.1016/S0065-3233(03)66002-X).
- [167] Verlet L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys Rev* 1967;159:98–103. <https://doi.org/10.1103/PhysRev.159.98>.
- [168] MacKerell AD, Bashford D, Bellott M, et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* 1998;102:3586–616. <https://doi.org/10.1021/jp973084f>.
- [169] Mulliken RS. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. II. Overlap Populations, Bond Orders, and Covalent Bond Energies. *J Chem Phys* 1955;23:1841–6. <https://doi.org/10.1063/1.1740589>.
- [170] Leach AR. *Molecular modelling: principles and applications*. Harlow, England : Longman; 1996.
- [171] Frenkel D, Smit B. *Understanding Molecular Simulation*. ScienceDirect 2002. <http://www.sciencedirect.com:5070/book/monograph/9780122673511/understanding-molecular-simulation> (accessed February 28, 2026).
- [172] Allen MP, Tildesley DJ. *Computer simulation of liquids*. 2nd ed. Oxford: Oxford university press; 2017.
- [173] Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J Chem Phys* 1993;98:10089–92. <https://doi.org/10.1063/1.464397>.
- [174] Jorgensen WL, Chandrasekhar J, Madura JD, et al. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–35. <https://doi.org/10.1063/1.445869>.
- [175] Kollman PA, Massova I, Reyes C, et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc Chem Res* 2000;33:889–97. <https://doi.org/10.1021/ar000033j>.
- [176] Srinivasan J, Cheatham TE, Cieplak P, et al. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J Am Chem Soc* 1998;120:9401–9. <https://doi.org/10.1021/ja981844+>.
- [177] Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* 2004;55:383–94. <https://doi.org/10.1002/prot.20033>.

- [178] Sitkoff D, Sharp KA, Honig B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J Phys Chem* 1994;98:1978–88. <https://doi.org/10.1021/j100058a043>.
- [179] Miller BR, McGee TD, Swails JM, et al. MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput* 2012;8:3314–21. <https://doi.org/10.1021/ct300418h>.
- [180] Case DA. Normal mode analysis of protein dynamics. *Current Opinion in Structural Biology* 1994;4:285–90. [https://doi.org/10.1016/S0959-440X\(94\)90321-2](https://doi.org/10.1016/S0959-440X(94)90321-2).
- [181] Andricioaei I, Karplus M. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J Chem Phys* 2001;115:6289–92. <https://doi.org/10.1063/1.1401821>.
- [182] Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* 2013;9:3084–95. <https://doi.org/10.1021/ct400341p>.
- [183] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery* 2015;10:449–61. <https://doi.org/10.1517/17460441.2015.1032936>.
- [184] Hou T, Wang J, Li Y, et al. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J Chem Inf Model* 2011;51:69–82. <https://doi.org/10.1021/ci100275a>.
- [185] Wang E, Sun H, Wang J, et al. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem Rev* 2019;119:9478–508. <https://doi.org/10.1021/acs.chemrev.9b00055>.
- [186] Greenidge PA, Kramer C, Mozziconacci J-C, et al. MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement. *J Chem Inf Model* 2013;53:201–9. <https://doi.org/10.1021/ci300425v>.
- [187] Gohlke H, Kiel C, Case DA. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *Journal of Molecular Biology* 2003;330:891–913. [https://doi.org/10.1016/S0022-2836\(03\)00610-7](https://doi.org/10.1016/S0022-2836(03)00610-7).
- [188] Hopkins AL, Groom CR, Alex A. Ligand efficiency: a useful metric for lead selection. *Drug Discovery Today* 2004;9:430–1. [https://doi.org/10.1016/S1359-6446\(04\)03069-7](https://doi.org/10.1016/S1359-6446(04)03069-7).
- [189] Cavalluzzi MM, Mangiatordi GF, Nicolotti O, et al. Ligand efficiency metrics in drug discovery: the pros and cons from a practical perspective.

- Expert Opinion on Drug Discovery 2017;12:1087–104.
<https://doi.org/10.1080/17460441.2017.1365056>.
- [190] Abad-Zapatero C, Metz JT. Ligand efficiency indices as guideposts for drug discovery. *Drug Discovery Today* 2005;10:464–9.
[https://doi.org/10.1016/S1359-6446\(05\)03386-6](https://doi.org/10.1016/S1359-6446(05)03386-6).
- [191] Hann MM. Molecular obesity, potency and other addictions in drug discovery. *Med Chem Commun* 2011;2:349–55.
<https://doi.org/10.1039/C1MD00017A>.
- [192] Mongan J, Simmerling C, McCammon JA, et al. Generalized Born Model with a Simple, Robust Molecular Volume Correction. *J Chem Theory Comput* 2007;3:156–69. <https://doi.org/10.1021/ct600085e>.
- [193] Nguyen H, Roe DR, Simmerling C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J Chem Theory Comput* 2013;9:2020–34. <https://doi.org/10.1021/ct3010485>.
- [194] Khaliha S, Bianchi A, Kovtun A, et al. Graphene oxide nanosheets for drinking water purification by tandem adsorption and microfiltration. *Separation and Purification Technology* 2022;300:121826.
<https://doi.org/10.1016/j.seppur.2022.121826>.
- [195] Trifoglio A, Mantovani S, Khaliha S, et al. Tailoring graphene oxide nanosheets by alkyl amine grafting for enhanced adsorption of PFASs in drinking water: a combined theoretical and experimental study. *Nanoscale* 2025;17:12124–33. <https://doi.org/10.1039/D5NR00502G>.
- [196] Hawkins GD, Cramer CJ, Truhlar DG. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J Phys Chem* 1996;100:19824–39. <https://doi.org/10.1021/jp961710n>.
- [197] Mantovani S, Khaliha S, Marforio TD, et al. Facile high-yield synthesis and purification of lysine-modified graphene oxide for enhanced drinking water purification. *Chem Commun* 2022;58:9766–9.
<https://doi.org/10.1039/D2CC03256B>.
- [198] Li D, Müller MB, Gilje S, et al. Processable aqueous dispersions of graphene nanosheets. *Nature Nanotech* 2008;3:101–5.
<https://doi.org/10.1038/nnano.2007.451>.
- [199] Dreyer DR, Park S, Bielawski CW, et al. The chemistry of graphene oxide. *Chem Soc Rev* 2009;39:228–40.
<https://doi.org/10.1039/B917103G>.
- [200] Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as

- regards priority substances in the field of water policyText with EEA relevance n.d.
- [201] Kovtun A, Bianchi A, Zambianchi M, et al. Core-shell graphene oxide-polymer hollow fibers as water filters with enhanced performance and selectivity. *Faraday Discuss* 2021;227:274–90. <https://doi.org/10.1039/C9FD00117D>.
 - [202] Tunioli F, Khaliha S, Mantovani S, et al. Adsorption of emerging contaminants by graphene related materials and their alginate composite hydrogels. *Journal of Environmental Chemical Engineering* 2023;11:109566. <https://doi.org/10.1016/j.jece.2023.109566>.
 - [203] Mantovani S, Marforio TD, Khaliha S, et al. Amino acid-driven adsorption of emerging contaminants in water by modified graphene oxide nanosheets. *Environ Sci: Water Res Technol* 2023;9:1030–40. <https://doi.org/10.1039/D2EW00871H>.
 - [204] Staples CA, Peterson DR, Parkerton TF, et al. The environmental fate of phthalate esters: A literature review. *Chemosphere* 1997;35:667–749. [https://doi.org/10.1016/S0045-6535\(97\)00195-1](https://doi.org/10.1016/S0045-6535(97)00195-1).
 - [205] Halden RU. Plastics and Health Risks. *Annual Review of Public Health* 2010;31:179–94. <https://doi.org/10.1146/annurev.publhealth.012809.103714>.
 - [206] Oehlmann J, Schulte-Oehlmann U, Kloas W, et al. A critical analysis of the biological impacts of plasticizers on wildlife. *Philos Trans R Soc Lond B Biol Sci* 2009;364:2047–62. <https://doi.org/10.1098/rstb.2008.0242>.
 - [207] Gao D-W, Wen Z-D. Phthalate esters in the environment: A critical review of their occurrence, biodegradation, and removal during wastewater treatment processes. *Science of The Total Environment* 2016;541:986–1001. <https://doi.org/10.1016/j.scitotenv.2015.09.148>.
 - [208] Gray LE Jr, Ostby J, Furr J, et al. Perinatal Exposure to the Phthalates DEHP, BBP, and DINP, but Not DEP, DMP, or DOTP, Alters Sexual Differentiation of the Male Rat. *Toxicol Sci* 2000;58:350–65. <https://doi.org/10.1093/toxsci/58.2.350>.
 - [209] Hlisníková H, Petrovičová I, Kolena B, et al. Effects and Mechanisms of Phthalates' Action on Reproductive Processes and Reproductive Health: A Literature Review. *International Journal of Environmental Research and Public Health* 2020;17:6811. <https://doi.org/10.3390/ijerph17186811>.
 - [210] Corton JC, Lapinskas PJ. Peroxisome Proliferator-Activated Receptors: Mediators of Phthalate Ester-Induced Effects in the Male Reproductive Tract? *Toxicol Sci* 2005;83:4–17. <https://doi.org/10.1093/toxsci/kfi011>.

- [211] Ward JM, Peters JM, Perella CM, et al. Receptor and Nonreceptor-Mediated Organ-Specific Toxicity of Di(2-ethylhexyl)phthalate (DEHP) in Peroxisome Proliferator-Activated Receptor α -Null Mice. *Toxicol Pathol* 1998;26:240–6. <https://doi.org/10.1177/019262339802600208>.
- [212] IARC. Some Chemicals Present in Industrial and Consumer Products, Food and Drinking-water. n.d.
- [213] Evaluation of Carcinogenic Risks to Humans: some industrial chemicals. 15-22 February 2000, Lyon, France. IARC Monogr Eval Carcinog Risks Hum 2000;77:1–529.
- [214] Singh S, Li SS-L. Epigenetic Effects of Environmental Chemicals Bisphenol A and Phthalates. *International Journal of Molecular Sciences* 2012;13:10143–53. <https://doi.org/10.3390/ijms130810143>.
- [215] Rowdhwal SSS, Chen J. Toxic Effects of Di-2-ethylhexyl Phthalate: An Overview. *BioMed Research International* 2018;2018:1750368. <https://doi.org/10.1155/2018/1750368>.
- [216] Consolidated TEXT: 32006R1907 — EN — 23.10.2025 n.d. <https://eur-lex.europa.eu/eli/reg/2006/1907/2025-10-23/eng> (accessed February 28, 2026).
- [217] Istvan ES, Deisenhofer J. Structural Mechanism for Statin Inhibition of HMG-CoA Reductase. *Science* 2001;292:1160–4. <https://doi.org/10.1126/science.1059344>.
- [218] Lennernäs H. Clinical Pharmacokinetics of Atorvastatin. *Clin Pharmacokinet* 2003;42:1141–60. <https://doi.org/10.2165/00003088-200342130-00005>.
- [219] Kasprzyk-Hordern B, Dinsdale RM, Guwy AJ. The occurrence of pharmaceuticals, personal care products, endocrine disruptors and illicit drugs in surface water in South Wales, UK. *Water Research* 2008;42:3498–518. <https://doi.org/10.1016/j.watres.2008.04.026>.
- [220] Aus Der Beek T, Weber F-A, Bergmann A, et al. Pharmaceuticals in the environment—Global occurrences and perspectives. *Environmental Toxicology and Chemistry* 2015;35:823–35. <https://doi.org/10.1002/etc.3339>.
- [221] Fent K, Weston AA, Caminada D. Ecotoxicology of human pharmaceuticals. *Aquatic Toxicology* 2006;76:122–59. <https://doi.org/10.1016/j.aquatox.2005.09.009>.
- [222] Wang Y, Hou Y, He C, et al. Toxic effects of acute and chronic atorvastatin exposure on antioxidant systems, autophagy processes, energy metabolism and life history in *Daphnia magna*. *Chemosphere* 2024;369:143792. <https://doi.org/10.1016/j.chemosphere.2024.143792>.

- [223] Richardson SD, Ternes TA. Water Analysis: Emerging Contaminants and Current Issues. *Anal Chem* 2018;90:398–428. <https://doi.org/10.1021/acs.analchem.7b04577>.
- [224] Mackay D, Shiu W-Y, Shiu W-Y, et al. Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals. 2nd ed. Boca Raton: CRC Press; 2006. <https://doi.org/10.1201/9781420044393>.
- [225] Giesy JP, Solomon KR, Cutler GC, et al. Ecological Risk Assessment of the Uses of the Organophosphorus Insecticide Chlorpyrifos, in the United States. In: Giesy JP, Solomon KR, editors. Ecological Risk Assessment for Chlorpyrifos in Terrestrial and Aquatic Systems in the United States, Cham: Springer International Publishing; 2014, p. 1–11. https://doi.org/10.1007/978-3-319-03865-0_1.
- [226] Solomon KR, Giesy JP, Kendall RJ, et al. Chlorpyrifos: Ecotoxicological Risk Assessment for Birds and Mammals in Corn Agroecosystems. *Human and Ecological Risk Assessment: An International Journal* 2001;7:497–632. <https://doi.org/10.1080/20018091094510>.
- [227] Eaton DL, Daroff RB, Autrup H, et al. Review of the Toxicology of Chlorpyrifos With an Emphasis on Human Exposure and Neurodevelopment. *Critical Reviews in Toxicology* 2008;38:1–125. <https://doi.org/10.1080/10408440802272158>.
- [228] Costa LG. Current issues in organophosphate toxicology. *Clinica Chimica Acta* 2006;366:1–13. <https://doi.org/10.1016/j.cca.2005.10.008>.
- [229] Slotkin TA. Cholinergic systems in brain development and disruption by neurotoxicants: nicotine, environmental tobacco smoke, organophosphates. *Toxicology and Applied Pharmacology* 2004;198:132–51. <https://doi.org/10.1016/j.taap.2003.06.001>.
- [230] SEVEN-YEAR NEURODEVELOPMENTAL SCORES AND PRENATAL EXPOSURE TO CHLORPYRIFOS, A COMMON AGRICULTURAL PESTICIDE. Environmental Hazards and Neurodevelopment, Apple Academic Press; 2015.
- [231] Commission Implementing Regulation (EU) 2020/17 of 10 January 2020 concerning the non-renewal of the approval of the active substance chlorpyrifos-methyl, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and of the Council concerning the placing of plant protection products on the market, and amending the Annex to Commission Implementing Regulation (EU) No 540/2011 (Text with EEA relevance). vol. 007. 2020.
- [232] Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as

- regards priority substances in the field of water policy Text with EEA relevance. vol. 226. 2013.
- [233] Yang K, Xing B. Adsorption of Organic Compounds by Carbon Nanomaterials in Aqueous Phase: Polanyi Theory and Its Application. *Chem Rev* 2010;110:5989–6008. <https://doi.org/10.1021/cr100059s>.
 - [234] Schniepp HC, Li J-L, McAllister MJ, et al. Functionalized Single Graphene Sheets Derived from Splitting Graphite Oxide. *J Phys Chem B* 2006;110:8535–9. <https://doi.org/10.1021/jp060936f>.
 - [235] Kruskal JB. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 1964;29:1–27. <https://doi.org/10.1007/BF02289565>.
 - [236] Borg I, Groenen PJF. Modern Multidimensional Scaling: Theory and Applications. Springer Science & Business Media; 2005.
 - [237] Ester M, Kriegel H-P, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise n.d.
 - [238] Tanimoto TT. An Elementary Mathematical Theory of Classification and Prediction. International Business Machines Corporation; 1958.
 - [239] Maggiora G, Vogt M, Stumpfe D, et al. Molecular Similarity in Medicinal Chemistry. *J Med Chem* 2014;57:3186–204. <https://doi.org/10.1021/jm401411z>.
 - [240] Chang C-E, Chen W, Gilson MK. Evaluating the Accuracy of the Quasiharmonic Approximation. *J Chem Theory Comput* 2005;1:1017–28. <https://doi.org/10.1021/ct0500904>.
 - [241] Hess B. Convergence of sampling in protein simulations. *Phys Rev E* 2002;65:031910. <https://doi.org/10.1103/PhysRevE.65.031910>.
 - [242] Laio A, Parrinello M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* 2002;99:12562–6. <https://doi.org/10.1073/pnas.202427399>.
 - [243] Sugita Y, Okamoto Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* 1999;314:141–51. [https://doi.org/10.1016/S0009-2614\(99\)01123-9](https://doi.org/10.1016/S0009-2614(99)01123-9).
 - [244] Bussi G, Laio A. Using metadynamics to explore complex free-energy landscapes. *Nat Rev Phys* 2020;2:200–12. <https://doi.org/10.1038/s42254-020-0153-0>.
 - [245] Calza A, Calvaresi M, Zerbetto F, et al. Martini 3 Coarse-Grain Model For Linear Perfluoroalkyl Substances. *J Chem Theory Comput* 2026:acs.jctc.6c00250. <https://doi.org/10.1021/acs.jctc.6c00250>.

- [246] Pablo JJ de. Coarse-Grained Simulations of Macromolecules: From DNA to Nanocomposites. *Annual Review of Physical Chemistry* 2011;62:555–74. <https://doi.org/10.1146/annurev-physchem-032210-103458>.
- [247] Saunders MG, Voth GA. Coarse-Graining Methods for Computational Biology. *Annual Review of Biophysics* 2013;42:73–93. <https://doi.org/10.1146/annurev-biophys-083012-130348>.
- [248] Klein ML, Shinoda W. Large-Scale Molecular Dynamics Simulations of Self-Assembling Systems. *Science* 2008;321:798–800. <https://doi.org/10.1126/science.1157834>.
- [249] Afify ND, Ferreiro-Rangel CA, Sweatman MB. Molecular Dynamics Investigation of Giant Clustering in Small-Molecule Solutions: The Case of Aqueous PEHA. *J Phys Chem B* 2022;126:8882–91. <https://doi.org/10.1021/acs.jpcb.2c04489>.
- [250] Peter C, Kremer K. Multiscale simulation of soft matter systems – from the atomistic to the coarse-grained level and back. *Soft Matter* 2009;5:4357. <https://doi.org/10.1039/b912027k>.
- [251] Noid WG. Perspective: Coarse-grained models for biomolecular systems. *The Journal of Chemical Physics* 2013;139:090901. <https://doi.org/10.1063/1.4818908>.
- [252] Ingólfsson HI, Lopez CA, Uusitalo JJ, et al. The power of coarse graining in biomolecular simulations. *WIREs Computational Molecular Science* 2014;4:225–48. <https://doi.org/10.1002/wcms.1169>.
- [253] Brosz M, Buck J, Grünewald F, et al. Coarse-grained Martini 3 model for collagen fibrils. *Biophysical Journal* 2025:S0006349525006630. <https://doi.org/10.1016/j.bpj.2025.10.012>.
- [254] Yangaliev D, Ozkan SB. Coarse-grained RNA model for the Martini 3 force field. *Biophysical Journal* 2025;0. <https://doi.org/10.1016/j.bpj.2025.07.034>.
- [255] Dettmann LF, Kühn O, Ahmed AA. Coarse-Grained Simulations of the Nanoplastic Interaction with Soil Organic Matter. *J Phys Chem B* 2025;129:11605–15. <https://doi.org/10.1021/acs.jpcb.5c06072>.
- [256] Muratov AD (Александр ДМ, Avetisov VA (Владик АА. Martini 3 application for the design of bistable nanomachines. *J Chem Phys* 2025;163:144902. <https://doi.org/10.1063/5.0295268>.
- [257] Grünewald F, Punt MH, Jefferys EE, et al. Martini 3 Coarse-Grained Force Field for Carbohydrates. *J Chem Theory Comput* 2022;18:7555–69. <https://doi.org/10.1021/acs.jctc.2c00757>.

- [258] Cambiaso S, Razmazma H, Shrestha R, et al. Martini 3 Coarse-Grained Model for Chitosan with Tunable Acetylation. *Biomacromolecules* 2025;26:7319–30. <https://doi.org/10.1021/acs.biomac.5c00683>.
- [259] Shrestha R, Alessandri R, Vögele M, et al. Martini 3 Coarse-Grained Models for Carbon Nanomaterials. *J Chem Theory Comput* 2025;21:9035–53. <https://doi.org/10.1021/acs.jctc.5c00923>.
- [260] Monticelli L, Kandasamy SK, Periole X, et al. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput* 2008;4:819–34. <https://doi.org/10.1021/ct700324x>.
- [261] Marrink SJ, Risselada HJ, Yefimov S, et al. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J Phys Chem B* 2007;111:7812–24. <https://doi.org/10.1021/jp071097f>.
- [262] Pedersen KB, Ingólfsson HI, Ramirez-Echemendia DP, et al. The Martini 3 Lipidome: Expanded and Refined Parameters Improve Lipid Phase Behavior. *ACS Cent Sci* 2025;11:1598–610. <https://doi.org/10.1021/acscentsci.5c00755>.
- [263] Souza PCT, Alessandri R, Barnoud J, et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat Methods* 2021;18:382–8. <https://doi.org/10.1038/s41592-021-01098-3>.
- [264] Marrink SJ, Monticelli L, Melo MN, et al. Two decades of Martini: Better beads, broader scope. *WIREs Computational Molecular Science* 2023;13:e1620. <https://doi.org/10.1002/wcms.1620>.
- [265] Alessandri R, Barnoud J, Gertsen AS, et al. Martini 3 Coarse-Grained Force Field: Small Molecules. *Advcd Theory and Sims* 2022;5:2100391. <https://doi.org/10.1002/adts.202100391>.
- [266] Murillo-Gelvez J, Dmitrenko O, Torralba-Sanchez TL, et al. pKa prediction of per- and polyfluoroalkyl acids in water using in silico gas phase stretching vibrational frequencies and infrared intensities. *Phys Chem Chem Phys* 2023;25:24745–60. <https://doi.org/10.1039/D3CP01390A>.
- [267] Balasuryia D, Queral-Beltran A, Vick T, et al. Experimental Determination of p Ka for 10 PFAS, Mono-, Di-, and Trifluoroacetic Acid by ¹⁹F-NMR. *Environ Sci Technol Lett* 2025;12:1238–46. <https://doi.org/10.1021/acs.estlett.5c00688>.
- [268] Rayne S, Forest K. Theoretical studies on the pK_a values of perfluoroalkyl carboxylic acids. *Nat Prec* 2010. <https://doi.org/10.1038/npre.2010.3829.2>.

- [269] Shirts MR, Klein C, Swails JM, et al. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J Comput Aided Mol Des* 2017;31:147–61. <https://doi.org/10.1007/s10822-016-9977-1>.
- [270] Abraham MJ, Murtola T, Schulz R, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1–2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- [271] Van Der Spoel D, Lindahl E, Hess B, et al. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* 2005;26:1701–18. <https://doi.org/10.1002/jcc.20291>.
- [272] BEKKER H, BERENDSEN H, DIJKSTRA E, et al. GROMACS - A PARALLEL COMPUTER FOR MOLECULAR-DYNAMICS SIMULATIONS: 4th International Conference on Computational Physics (PC 92). *PHYSICS COMPUTING '92* 1993:252–6.
- [273] Barnoud, Jonathan. CGBuilder 2021.
- [274] Leung SCE, Wanninayake D, Chen D, et al. Physicochemical properties and interactions of perfluoroalkyl substances (PFAS) - Challenges and opportunities in sensing and remediation. *Science of The Total Environment* 2023;905:166764. <https://doi.org/10.1016/j.scitotenv.2023.166764>.
- [275] Golden WG, Brown EM, Solem SE, et al. Complete conformational analyses of perfluoro-n-pentane, perfluoro-n-hexane, and perfluoro-n-heptane. *Journal of Molecular Structure: THEOCHEM* 2008;867:22–7. <https://doi.org/10.1016/j.theochem.2008.07.011>.
- [276] Mifkovic M, Van Hoomissen DJ, Vyas S. Conformational distributions of helical perfluoroalkyl substances and impacts on stability. *J Comput Chem* 2022;43:1656–61. <https://doi.org/10.1002/jcc.26967>.
- [277] Schilberg RN, Wei S, Twagirayezu S, et al. Conformational dynamics of perfluorooctanoic acid (PFOA) studied by molecular rotational resonance (MRR) spectroscopy. *Chemical Physics Letters* 2021;778:138789. <https://doi.org/10.1016/j.cplett.2021.138789>.
- [278] De Jong DH, Baoukina S, Ingólfsson HI, et al. Martini straight: Boosting performance using a shorter cutoff and GPUs. *Computer Physics Communications* 2016;199:1–7. <https://doi.org/10.1016/j.cpc.2015.09.014>.
- [279] Rowland RS, Taylor R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. *J Phys Chem* 1996;100:7384–91. <https://doi.org/10.1021/jp953141+>.

- [280] Rayne S, Forest K. Perfluoroalkyl sulfonic and carboxylic acids: A critical review of physicochemical properties, levels and patterns in waters and wastewaters, and treatment methods. *Journal of Environmental Science and Health, Part A* 2009;44:1145–99. <https://doi.org/10.1080/10934520903139811>.
- [281] Rayne S, Forest K. A comparative assessment of octanol-water partitioning and distribution constant estimation methods for perfluoroalkyl carboxylates and sulfonates. *Nat Prec* 2009. <https://doi.org/10.1038/npre.2009.3282.2>.
- [282] Meng P, Deng S, Du Z, et al. Effect of hydro-oleophobic perfluorocarbon chain on interfacial behavior and mechanism of perfluorooctane sulfonate in oil-water mixture. *Sci Rep* 2017;7:44694. <https://doi.org/10.1038/srep44694>.
- [283] Weiss J, de Boer J, Berger U, et al. PFAS analysis in water for the Global Monitoring Plan of the Stockholm Convention n.d.
- [284] Jing P, Rodgers PJ, Amemiya S. High Lipophilicity of Perfluoroalkyl Carboxylate and Sulfonate: Implications for Their Membrane Permeability. *J Am Chem Soc* 2009;131:2290–6. <https://doi.org/10.1021/ja807961s>.
- [285] Hu J, Yang X, Song X, et al. Elucidating environmental fate and toxicological mechanisms of ultrashort- and short-chain PFAS: Integrating machine learning, molecular modeling, and experimental validation. *Journal of Environmental Management* 2025;390:126277. <https://doi.org/10.1016/j.jenvman.2025.126277>.
- [286] Torralba-Sanchez TL, Di Toro DM, Dmitrenko O, et al. Modeling the Partitioning of Anionic Carboxylic and Perfluoroalkyl Carboxylic and Sulfonic Acids to Octanol and Membrane Lipid. *Environmental Toxicology and Chemistry* 2023;42:2317–28. <https://doi.org/10.1002/etc.5716>.
- [287] Minervino A, Belfield KD. Review of Recent Computational Research on the Adsorption of PFASs with a Variety of Substrates. *IJMS* 2024;25:3445. <https://doi.org/10.3390/ijms25063445>.
- [288] Du Z, Deng S, Bei Y, et al. Adsorption behavior and mechanism of perfluorinated compounds on various adsorbents—A review. *Journal of Hazardous Materials* 2014;274:443–54. <https://doi.org/10.1016/j.jhazmat.2014.04.038>.
- [289] Bresnahan CG, Schutt TC, Shukla MK. Exploration of functionalizing graphene and the subsequent impact on PFAS adsorption capabilities

- via molecular dynamics. *Chemosphere* 2023;345:140462. <https://doi.org/10.1016/j.chemosphere.2023.140462>.
- [290] Lamb BG, Ma B. PFAS self-assembly and adsorption dynamics on graphene: molecular insights into chemical and environmental influences. *Nanoscale* 2025;17:10632–43. <https://doi.org/10.1039/D4NR04995K>.
- [291] Costanza J, Arshadi M, Abriola LM, et al. Accumulation of PFOA and PFOS at the Air–Water Interface. *Environ Sci Technol Lett* 2019;6:487–91. <https://doi.org/10.1021/acs.estlett.9b00355>.
- [292] Viberg H, Eriksson P. Perfluorooctane sulfonate (PFOS) and perfluorooctanoic acid (PFOA). *Reproductive and Developmental Toxicology*, Elsevier; 2011, p. 623–35. <https://doi.org/10.1016/B978-0-12-382032-7.10047-5>.
- [293] Wang X, Zhang H, Ham S, et al. Graphene Oxide and Its Derivatives as Adsorbents for PFOA Molecules. *J Phys Chem B* 2023;127:9620–9. <https://doi.org/10.1021/acs.jpcc.3c04762>.
- [294] Cambiaso S, Tagliabue A, Bochicchio D, et al. Solid–Liquid Interfaces in MARTINI 3: Modeling Approaches and the Case Study of Silica Surfaces in Water 2025. <https://doi.org/10.26434/chemrxiv-2024-bkmwz-v2>.
- [295] Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc* 1996;118:11225–36. <https://doi.org/10.1021/ja9621760>.
- [296] Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* 2020;17:261–72. <https://doi.org/10.1038/s41592-019-0686-2>.
- [297] Gowers R, Linke M, Barnoud J, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations, Austin, Texas: 2016, p. 98–105. <https://doi.org/10.25080/Majora-629e541a-00e>.
- [298] Impey RW, Madden PA, McDonald IR. Hydration and mobility of ions in solution. *J Phys Chem* 1983;87:5071–83. <https://doi.org/10.1021/j150643a008>.
- [299] Luzar A, Chandler D. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Phys Rev Lett* 1996;76:928–31. <https://doi.org/10.1103/PhysRevLett.76.928>.
- [300] Yushmanov PV, Furó I, Stilbs P. Micellar Kinetics of a Fluorosurfactant through Stopped-Flow NMR. *Langmuir* 2006;22:2002–4. <https://doi.org/10.1021/la0529660>.
- [301] Aniansson EAG, Wall SN. Kinetics of step-wise micelle association. *J Phys Chem* 1974;78:1024–30. <https://doi.org/10.1021/j100603a016>.

- [302] Barraud E, Dalmazzone C, Mouret A, et al. A Coarse-Grained Model Describing the Critical Micelle Concentration of Perfluoroalkyl Surfactants in Ionic Aqueous Phase. *Langmuir* 2025;41:7272–82. <https://doi.org/10.1021/acs.langmuir.4c04293>.
- [303] López-Fontán JL, Sarmiento F, Schulz PC. The aggregation of sodium perfluorooctanoate in water. *Colloid Polym Sci* 2005;283:862–71. <https://doi.org/10.1007/s00396-004-1228-7>.
- [304] Kancharla S, Canales E, Alexandridis P. Perfluorooctanoate in Aqueous Urea Solutions: Micelle Formation, Structure, and Microenvironment. *International Journal of Molecular Sciences* 2019;20:5761. <https://doi.org/10.3390/ijms20225761>.
- [305] Klevan C, Caines S, Gomes A, et al. Accurate Determination of Perfluorooctanoate Aqueous Solubility, Critical Micelle Concentration, and Acid Dissociation Constant. *Environ Sci Technol Lett* 2024;11:1398–405. <https://doi.org/10.1021/acs.estlett.4c00858>.
- [306] Kancharla S, Dong D, Bedrov D, et al. Structure and Interactions in Perfluorooctanoate Micellar Solutions Revealed by Small-Angle Neutron Scattering and Molecular Dynamics Simulations Studies: Effect of Urea. *Langmuir* 2021;37:5339–47. <https://doi.org/10.1021/acs.langmuir.1c00433>.
- [307] Dong D, Kancharla S, Hooper J, et al. Controlling the self-assembly of perfluorinated surfactants in aqueous environments. *Phys Chem Chem Phys* 2021;23:10029–39. <https://doi.org/10.1039/D1CP00049G>.
- [308] Kancharla S, Jahan R, Bedrov D, et al. Role of chain length and electrolyte on the micellization of anionic fluorinated surfactants in water. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 2021;628:127313. <https://doi.org/10.1016/j.colsurfa.2021.127313>.
- [309] González-Pérez A, Ruso JM, Prieto G, et al. The self-aggregation of sodium perfluorooctanoate in aqueous solution at different temperatures. *Journal of Surfactants and Detergents* 2004;7:387–95. <https://doi.org/10.1007/s11743-004-0323-9>.
- [310] Mukerjee P, Korematsu K, Okawauchi M, et al. Effect of temperature on the electrical conductivity and the thermodynamics of micelle formation of sodium perfluorooctanoate. *J Phys Chem* 1985;89:5308–12. <https://doi.org/10.1021/j100270a037>.
- [311] Hill C, Czajka A, Hazell G, et al. Surface and bulk properties of surfactants used in fire-fighting. *Journal of Colloid and Interface Science* 2018;530:686–94. <https://doi.org/10.1016/j.jcis.2018.07.023>.

- [312] Ohshima H. Surface Charge Density/Surface Potential Relationship for a Spherical Colloidal Particle in a Solution of General Electrolytes. *Journal of Colloid and Interface Science* 1995;171:525–7. <https://doi.org/10.1006/jcis.1995.1214>.
- [313] Ohshima H. *Theory of Colloid and Interfacial Electric Phenomena*. Elsevier; 2006.
- [314] Manning GS. Limiting Laws and Counterion Condensation in Polyelectrolyte Solutions I. Colligative Properties. *J Chem Phys* 1969;51:924–33. <https://doi.org/10.1063/1.1672157>.
- [315] Oosawa F. A simple theory of thermodynamic properties of polyelectrolyte solutions. *Journal of Polymer Science* 1957;23:421–30. <https://doi.org/10.1002/pol.1957.1202310335>.
- [316] Sanga N, Ahrens L, Rennie AR. Counterion influence on perfluorooctanoate micelle characteristics: Shape, size, and interactions. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 2025;726:137912. <https://doi.org/10.1016/j.colsurfa.2025.137912>.
- [317] Yan B, Alessandri R, Marrink SJ, et al. Insight into the Self-Assembly Behaviors of Per- and Polyfluoroalkyl Substances Using a “Computational Microscope.” *Environ Sci Technol Lett* 2025;12:626–31. <https://doi.org/10.1021/acs.estlett.4c01081>.
- [318] Giulieri F, Krafft M-P. Can the formation of vesicles from single-chain perfluoroalkylated amphiphiles be predicted? *Colloids and Surfaces A: Physicochemical and Engineering Aspects* 1994;84:121–7. [https://doi.org/10.1016/0927-7757\(93\)02732-T](https://doi.org/10.1016/0927-7757(93)02732-T).
- [319] Krafft M, Giuleiri F, Riess JG. Can Single-Chain Perfluoroalkylated Amphiphiles Alone Form Vesicles and Other Organized Supramolecular Systems? *Angew Chem Int Ed Engl* 1993;32:741–3. <https://doi.org/10.1002/anie.199307411>.
- [320] Rosholm KR, Arouri A, Hansen PL, et al. Characterization of Fluorinated Catansomes: A Promising Vector in Drug-Delivery. *Langmuir* 2012;28:2773–81. <https://doi.org/10.1021/la2039834>.
- [321] Chaban VV, Verspeek B, Khandelia H. Novel Ultrathin Membranes Composed of Organic Ions. *J Phys Chem Lett* 2013;4:1216–20. <https://doi.org/10.1021/jz400424f>.
- [322] Alam P, Khan AH, Islam R, et al. Identification of prevalent leachate percolation of municipal solid waste landfill: a case study in India. *Sci Rep* 2024;14:8910. <https://doi.org/10.1038/s41598-024-58693-5>.
- [323] Benskin JP, Li B, Ikonomou MG, et al. Per- and Polyfluoroalkyl Substances in Landfill Leachate: Patterns, Time Trends, and Sources.

- Environ Sci Technol 2012;46:11532–40.
<https://doi.org/10.1021/es302471n>.
- [324] Chen Y, Zhang H, Liu Y, et al. Evaluation of per- and polyfluoroalkyl substances (PFAS) in leachate, gas condensate, stormwater and groundwater at landfills. *Chemosphere* 2023;318:137903. <https://doi.org/10.1016/j.chemosphere.2023.137903>.
- [325] Robey NM, Liu Y, Tolaymat TM, et al. Per- and polyfluoroalkyl substances (PFAS) profiles in primary and secondary landfill leachates: Indications of transformation, liner interactions, and other PFAS sources. *Journal of Hazardous Materials* 2025;494:138705. <https://doi.org/10.1016/j.jhazmat.2025.138705>.
- [326] Gómez M, Corona F, Hidalgo MD. Variations in the properties of leachate according to landfill age. *Desalination and Water Treatment* 2019;159:24–31. <https://doi.org/10.5004/dwt.2019.24106>.
- [327] Saghi MH, Nadimi H, Eslami A, et al. Characteristics and pollution indices of leachates from municipal solid waste landfills in Iranian metropolises and their implications for MSW management. *Sci Rep* 2024;14:27285. <https://doi.org/10.1038/s41598-024-78630-w>.
- [328] Nika M-C, Alygizakis N, Arvaniti OS, et al. Non-target screening of emerging contaminants in landfills: A review. *Current Opinion in Environmental Science & Health* 2023;32:100430. <https://doi.org/10.1016/j.coesh.2022.100430>.
- [329] He X-S, Pan Q, Xi B-D, et al. Volatile and semi-volatile organic compounds in landfill leachate: Concurrence, removal and the influencing factors. *Water Research* 2023;245:120566. <https://doi.org/10.1016/j.watres.2023.120566>.
- [330] Propp VR, De Silva AO, Spencer C, et al. Organic contaminants of emerging concern in leachate of historic municipal landfills. *Environmental Pollution* 2021;276:116474. <https://doi.org/10.1016/j.envpol.2021.116474>.
- [331] Roy JW, Propp VR, Hua T, et al. Per- and poly-fluoroalkyl substances (PFAS) contamination of surface waters by historic landfills via groundwater plumes: ecosystem exposure and downstream mass loading. *Environ Sci: Processes Impacts* 2025;27:1074–87. <https://doi.org/10.1039/D4EM00612G>.
- [332] Euler T. Micropollutants in Urban Wastewater n.d.
- [333] Samy M, Gar Alalm M, Khalil MN, et al. Treatment of hazardous landfill leachate containing 1,4 dioxane by biochar-based photocatalysts in a

- solar photo-oxidation reactor. *Journal of Environmental Management* 2023;332:117402. <https://doi.org/10.1016/j.jenvman.2023.117402>.
- [334] Nishmitha PS, Akhilghosh KA, Aiswriya VP, et al. Understanding emerging contaminants in water and wastewater: A comprehensive review on detection, impacts, and solutions. *Journal of Hazardous Materials Advances* 2025;18:100755. <https://doi.org/10.1016/j.hazadv.2025.100755>.
- [335] Das S, Parida VK, Tiwary CS, et al. Emerging Contaminants in the Aquatic Environment: Fate, Occurrence, Impacts, and Toxicity. In: Shah MP, Vasantharaj K, Saranya N, editors. *ACS Symposium Series*, vol. 1475, Washington, DC: American Chemical Society; 2024, p. 1–32. <https://doi.org/10.1021/bk-2024-1475.ch001>.
- [336] Li X, Shen X, Jiang W, et al. Comprehensive review of emerging contaminants: Detection technologies, environmental impact, and management strategies. *Ecotoxicology and Environmental Safety* 2024;278:116420. <https://doi.org/10.1016/j.ecoenv.2024.116420>.
- [337] Hilpert C, Beranger L, Souza PCT, et al. Facilitating CG Simulations with MAD: The MArtini Database Server. *J Chem Inf Model* 2023;63:702–10. <https://doi.org/10.1021/acs.jcim.2c01375>.
- [338] Tunioli F, Khaliha S, Mantovani S, et al. Adsorption of emerging contaminants by graphene related materials and their alginate composite hydrogels. *Journal of Environmental Chemical Engineering* 2023;11:109566. <https://doi.org/10.1016/j.jece.2023.109566>.
- [339] Kovtun A, Bianchi A, Zambianchi M, et al. Core–shell graphene oxide–polymer hollow fibers as water filters with enhanced performance and selectivity. *Faraday Discuss* 2021;227:274–90. <https://doi.org/10.1039/C9FD00117D>.