Fakultät für
Mathematik und
Naturwissenschaften

RWTH AACHEN UNIVERSITY

From Shapes to Signatures: How Molecules Become Barcodes for Data-Driven Chemistry

Von der Fakultät für Mathematik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

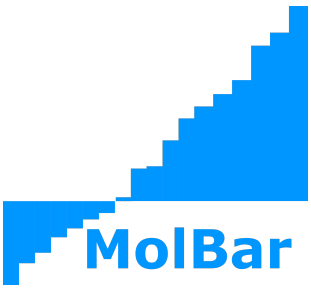
Nils van Staalduinen

aus

Höxter, Deutschland

Berichter:	Jun.-Prof. Dr. Christoph Bannwarth Univ.-Prof. Dr. Arne Lüchow
Tag der mündlichen Prüfung	19.08.2026

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.



Lernen wir träumen [...], dann finden wir vielleicht die Wahrheit.

— AUGUST KEKULÉ
11. März 1890
Benzolfest der Deutschen Chemischen Gesellschaft

Declaration of Authorship

I, Nils van Staalduinen, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research. I do solemnly swear that:

1. This work was done wholly or mainly while in candidature for the doctoral degree at this faculty and university;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this university or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others or myself, this is always clearly attributed;
4. Where I have quoted from the work of others or myself, the source is always given. This thesis is entirely my own work, with the exception of such quotations;
5. I have acknowledged all major sources of assistance;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published before as
    - **N. van Staalduinen** and C. Bannwarth, *Digit. Discov.*, 2024, 3, 2298–2319, DOI: `10.1039/D4DD00028A`, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.
    - **N. van Staalduinen** and C. Bannwarth, *J. Chem. Inf. Model.*, 2026, 66, 10935–10955, DOI: `10.1021/acs.jcim.6c00688`, with permission from the American Chemical Society.

Aachen, _____ _____

Name of Student

Declaration on the Use of AI Tools

Area of Application	AI Tool used?	Tool
Literature Research	Yes	Gemini 3.1, GPT-4o, GPT-5.1, Mistral Large 3
Spelling	Yes	DeepL Write, Gemini 3.1, Mistral Large 3
Phrasing/Paraphrasing	Yes	DeepL Write, Gemini 3.1, Mistral Large 3
Generation of Images	No	
Generation of Code	Yes	Claude Sonnet 4.5, Gemini 3.1, GPT-4o, GPT-5.1-Codex
Other	No	

The author used AI tools for literature research, which provided links to the original publications. These publications were accessed and read by the author to ensure a comprehensive understanding of the content, which were integrated into the thesis based on the author's own interpretations and conclusions. The author wrote this thesis based on his own ideas, thoughts, interpretations, and conclusions. Then, the author's text was checked for grammar, spelling, and wording using the aforementioned AI tools to improve clarity and readability. The author carefully reviewed, checked, and curated the proposals made by the model. The code was generated by the author, with the assistance of AI tools for debugging and optimization, but the intellectual content and structure of the code are solely the author's work. The model did not contribute to the intellectual content of the thesis, which is entirely the author's own work.

I, Nils van Staalduinen, declare the accuracy and completeness of the information provided in the table above:

Aachen, _____ _____

Name of Student

Publication List

- **N. van Staalduinen** and C. Bannwarth, *J. Chem. Inf. Model.*, 2026, 66, 10935–10955, DOI: 10.1021/acs.jcim.6c00688.
- C. Zurek, R. A. Mallaev, A. C. Paul, **N. van Staalduinen**, P. Pracht, R. Ellerbrock and C. Bannwarth, *J. Chem. Theory Comput.*, 2025, 21, 1459–1475, DOI: 10.1021/acs.jctc.4c01275.
- **N. van Staalduinen** and C. Bannwarth, *Digit. Discov.*, 2024, 3, 2298–2319, DOI: 10.1039/D4DD00028A.
- R. J. Kutta, J. Großkopf, **N. van Staalduinen**, A. Seitz, P. Pracht, S. Breitenlechner, C. Bannwarth, P. Nürnberger and T. Bach, *J. Am. Chem. Soc.*, 2023, 145, 2354–2363, DOI: 10.1021/jacs.2c11265.
- L. Reuter, **N. van Staalduinen**, J. Simons, J. Ludovicy and A. Lüchow, *J. Chem. Phys.*, 2022, 156, DOI: 10.1063/5.0090607.
- A. Ohligschläger, **N. van Staalduinen**, C. Cormann, J. Mühlhans, J. Wurm and M. A. Liauw, *Chem. Methods.*, 2021, 1, 181–191, DOI: 10.1002/cmtd.202000056.

Acknowledgments

First and foremost, I want to thank my supervisor, **Prof. Dr. rer. nat. Christoph Bannwarth**, for his continuous guidance and support over the last 4.5 years. Starting in a newly founded group with only Christoph, Pit, and myself, and watching it grow into a thriving team of more than a dozen researchers has been an incredible journey. Through Christoph's dedicated mentorship, I learned the importance of scientific rigor, sharpened my creativity, and developed essential skills in writing clean code and academic papers. I feel truly fortunate to have worked on such an inspiring and fascinating topic. I am incredibly grateful for his active support of my academic development beyond the university. His encouragement allowed me to participate in the BASF Summer Course 2023, travel to the ACS Fall Conference 2024 in Denver, and visit Stanford University. Moreover, I greatly appreciate his flexibility in allowing me to work at BASF alongside my PhD, giving me industry insights in cheminformatics that deeply enriched this thesis. These past years have shaped me both professionally and personally, and I owe a great deal of this growth to his steady support and trust.

I am also sincerely grateful to **Prof. Dr. rer. nat. Arne Lüchow** for taking on the role of second supervisor and for his valuable input and support throughout this thesis.

First, I want to thank **Pit** for making the start in 2021 in this small group so enjoyable. I am grateful to **Mike** for his dry humor, and the countless memes that made our time in the office so memorable. A special thanks to **Gereon** for the many fruitful scientific discussions and our shared running sessions as we prepared for the Hartree-Fock Cup. I want to thank **Ying** for being the best source for rumors about when Chic Town will open again. I also want to thank **Christopher** and **Martin** for organizing the "BRudi Rockt" and "Pubquiz" events—I hope to participate in them again in the future. I want to thank **Daria** and **Denise** for making the office a more welcoming place for everyone (ducks included). To **Mark** and **Hannah**, thank you for keeping the BeNeLux quota in the group alive even after Pit and I leave. I am especially thankful to **Dean** for making my visit to Stanford so enjoyable; it was wonderful to have you in Aachen for half a year as well. I look forward to visiting you in the U.S. in the future. I want to acknowledge my research students—**Janis**, **Christopher**, **Hannah**, and **Felix**—for their valuable scientific contributions.

I would also like to thank my friends for making the last 10 years in Aachen so wonderful. First of all, I would like to thank my group of friends with whom everything started here in

Aachen: **Anna**, **Timon**, **Justus**, **Maria**, and **Sabrina**. In particular, I want to thank **Anna** and **Timon**. It is amazing to have friends who have been there from the very beginning—starting as lab protocol buddies in the first semester, facing all the challenges of our Bachelor studies together, and remaining close friends after all these years. I would also like to thank **Ninon**; I am very happy that we have become such good friends, although we could have met four years earlier ;). The training sessions at the gym on Wednesdays were always a highlight of the week, even if I could hardly walk the next day (insert peach emoji here). Finally, I would like to thank my friends from my hometown—**Mario**, **Tim**, **Sebastian**, and **Christoph**—for the amazing times we shared in the past and the many vacation trips together. Especially, I want to thank **Mario** for being a great friend since our school days and for countless memorable trips to sporting events and groundhopping adventures.

Am meisten möchte ich aber **meinen Eltern** danken, die mich immer bei allem unterstützt haben, was ich mache, und mich immer ermutigt haben, meinen eigenen Weg zu gehen. Ich bin euch sehr dankbar für eure Unterstützung. Ohne euch wäre das alles nicht möglich gewesen. Danke, dass ihr mich auf meinem Weg begleitet habt und immer für mich da seid.

Für meine Eltern.

Abstract

In computational chemistry, 3D molecular structures are compared across various tasks, such as conformer sampling, reaction discovery, structural deduplication, and chemical database registration. Each of these tasks requires reliable molecular identifiers to accurately determine structural identity. The Simplified Molecular Input Line Entry Specification (SMILES) and the IUPAC International Chemical Identifier (InChI) are the most widely used identifiers for this purpose and are considered the gold standard in 2D cheminformatics. However, these identifiers were not originally designed for 3D molecular representations. They require external tools to convert 3D geometries into 2D formats, and different implementations may yield inconsistent results, precluding a strict one-to-one mapping between 3D structures and their identifiers. Additionally, their applicability is largely limited to ground-state organic molecules and standard transition-metal complexes. Finally, the stereochemical frameworks of SMILES and InChI, built around individual stereocenters and stereobonds, offer limited coverage of non-central chirality, such as helical or planar chirality.

To address these limitations, this work introduces the Molecular Barcode (MolBar), a molecular identifier purpose-built for quantum chemistry applications, departing fundamentally from the graph-based paradigm of existing identifiers. Rather than describing a molecule as a set of bonded atoms, MolBar encodes molecular identity through the 3D shape of a set of rigid substructures, each defined by its relative and absolute configuration. This approach naturally captures all non-local stereochemical configurations beyond individual stereocenters or stereobonds, covering the full periodic table and chirality types including central, axial, helical, and planar chirality. By deriving all information directly from 3D coordinates, MolBar eliminates the need for external tools to determine connectivity, bond orders, or stereochemistry—including the assignment of atropisomeric axes— and thereby guarantees a rigorous one-to-one mapping between a 3D geometry and its identifier string.

The underlying algorithms have been rigorously benchmarked against state-of-the-art tools, demonstrating exceptional robustness and computational efficiency. When deployed in large-scale 3D structure deduplication and quantum chemical space exploration workflows, MolBar exhibits superior performance compared to existing identifiers. These results demonstrate that MolBar is uniquely suited as a molecular identifier for modern quantum chemistry. MolBar is available as an open-source Python package via `pip install molbar`.

Contents

Acknowledgments **ix**

Abstract **xiii**

1 Introduction **1**

2 Theory **7**
2.1 Quantum Chemistry and Potential Energy Surface 7
2.2 Molecular Representations in Quantum Chemistry and Cheminformatics 15
2.3 Graph Theory in Cheminformatics 24

3 Motivation **29**

4 Theoretical Derivation of MolBar **37**
4.1 Topology: Molecular Connectivity 39
4.2 Topography: Relative Configuration 40
4.3 Absolute Configuration: Chirality 46

5 Practical Implementation **49**
5.1 Outline of this Chapter 51
5.2 Topology Matrix: Determining Connectivity from 3D Coordinates 53
5.2.1 Introduction 53
5.2.2 MolBar Topology Model 55
5.2.3 Dataset and Computational Details 56
5.2.4 Results and Discussion 57
5.3 Rigidity Matrix: Electronic Hindrance Classifier 61
5.3.1 Introduction 61
5.3.2 Electronic Hindrance Model 65
5.3.3 Dataset 68
5.3.4 Model Training and Evaluation 69
5.3.5 Results and Discussion 71
5.4 Rigidity Matrix: Steric Hindrance Model 80
5.4.1 Introduction 80
5.4.2 Steric Hindrance Model 82
5.4.3 Computational Protocol and Details 84
5.4.4 Validation of the Computational Protocol using Experimental Data . . . 86

5.4.5 Dataset . . . 89
5.4.6 Model Training and Evaluation . . . 89
5.4.7 Results and Discussion . . . 91
5.5 Topography Matrix: Structure Unification . . . 95
5.5.1 Introduction . . . 95
5.5.2 Structure Unification Procedure . . . 99
5.5.3 Dataset . . . 107
5.5.4 Computational Details . . . 108
5.5.5 Results and Discussion . . . 109
5.6 Absolute Configuration Matrix: Chirality Index and Priorities . . . 118
5.6.1 Determining Absolute Configuration of Arbitrary Fragments . . . 119
5.6.2 Assignment of Atomic and Fragment Priorities . . . 120
5.7 Structure of the Molecular Barcode (MolBar) Identifier . . . 123

6 Results and Discussion: Benchmarking MolBar and Application to Chemical Problems 127
6.1 Mapping the Potential Energy Surface through MolBar . . . 129
6.2 Representation of Stereoisomers beyond Atom-Centered Chirality . . . 134
6.3 Permutation Invariance Test . . . 139
6.4 Assessment of Computational Efficiency of MolBar . . . 141
6.5 Application of MolBar as a Molecular Representation in Databases and Database Deduplication . . . 145
6.6 Application of MolBar in Chemical Space Exploration . . . 151
6.6.1 Molecular Dynamics Simulation . . . 151
6.6.2 Conformer Sampling . . . 152
6.6.3 Reaction Discovery . . . 154

7 Conclusion and Outlook 161

Bibliography 164

List of Mathematical Symbols 183

Acronyms 187

A Figures A-1
A.1 Electronic Model: Cross-Validation Folds . . . A-1
A.2 Steric Model: Dataset . . . A-4
A.3 Steric Model: Cross-Validation Folds . . . A-19

B First and Second Order Derivatives of the Structure Unification Force Field A-23

1 Introduction

The conceptual framework for representing molecular structures was developed by Gilbert N. Lewis in the early 20th century. In his seminal work "The Atom and the Molecule", Lewis initially proposed that electrons occupy fixed positions at the corners of a cube.[1] From this model, the octet rule emerged: atoms tend to have eight electrons in their valence shell, often depicted as dots around the atomic symbol. Shared pairs of electrons forming covalent bonds were represented as pairs of dots between atoms, while lone pairs were shown as dots adjacent to a single atom. This led to the Lewis structure, a model that is still used today to depict molecular structures and understand chemical bonding.

However, this static, classical picture of chemical bonding was later supplemented by quantum mechanics. Niels Bohr's 1913 atomic model introduced quantized electron orbits while retaining certain classical elements. This model was able to explain the discrete spectral lines of hydrogen.[2] In 1924, Louis de Broglie proposed that particles exhibit wave-like properties, and introduced the concept of matter waves.[3] The development of quantum mechanics in the 1920s through Heisenberg's matrix mechanics[4] and Schrödinger's wave mechanics[5] replaced classical determinism with probabilistic wavefunctions. Werner Heisenberg's uncertainty principle further challenged classical notions by showing that position and momentum cannot be simultaneously measured with arbitrary precision.[6]

In 1927, Max Born and Robert J. Oppenheimer introduced an approximation to separate nuclear and electronic motion in molecules, laying the foundation for modern molecular quantum chemistry.[7] Building on these developments, Walter Heitler and Fritz London were the first to successfully apply quantum mechanical principles to the description of chemical bonding in the dihydrogen molecule.[8] Their work demonstrated that covalent bonding stems from the quantum mechanical superposition of electronic wave functions. These findings revealed that electron spin pairing and exchange symmetry play important roles in bond formation. This established the foundations of valence bond (VB) theory. Here, chemical bonding is described in terms of the interaction of atomic orbitals, which combine to form localized bonds. In the subsequent years, molecular orbital (MO) theory emerged as an alternative and complementary approach, initiated by Hund and Mulliken.[9] In this context, electrons in a molecule are not restricted to specific bonds between atom pairs, but are instead described by molecular orbitals that extend across the entire molecule and are formed under the influence of all atomic nuclei. In 1928, D. R. Hartree approximated the many-electron wavefunction as a product of one-electron wavefunctions, known as the Hartree product. This allowed him to obtain approximate solutions of the Schrödinger equation from first principles.[10] In this model, electrons are

treated as independent particles that interact with an average potential formed by the nuclei and other electrons. John Slater and Vladimir Fock subsequently refined Hartree's product *ansatz* by introducing the Slater determinant. This enforces the antisymmetry of the many-electron wavefunction and ensures compliance with the Pauli exclusion principle formulated in 1925.[11,12] This development culminated in the Hartree–Fock method, which accurately models exchange interactions but still neglects electron correlation beyond the mean-field approximation. In 1929, John Lennard-Jones proposed the linear combination of atomic orbitals (LCAO) method, which provided a description of bonding in second-period diatomic molecules such as N_2 and O_2.[9] This method represents molecular orbitals as linear combinations of atomic orbitals. In the 1930s, Erich Hückel developed a simplified LCAO approach focusing on π-electron systems in conjugated hydrocarbons. This work led to the formulation of Hückel's rule and significantly advanced the understanding of aromatic stability.[13]

An important development in computational quantum chemistry was the formulation of the Roothaan–Hall equations in 1951. This reformulation combined the Hartree–Fock method with the LCAO approach by expressing the molecular orbitals as linear combinations of a finite set of basis functions.[14,15] In its original form, the Hartree–Fock method is expressed as an integro-differential equation, which makes a direct numerical solution considerably more intricate. The Roothaan–Hall equations recast the Hartree–Fock problem as a matrix eigenvalue equation, thereby enabling its solution using standard linear algebra techniques. The use of linear algebra made the Hartree–Fock method suitable for implementation on emerging computational platforms. The rapid development of electronic computing technology during the 1940s provided essential technical support for these advances. This progress culminated in 1956 at the Massachusetts Institute of Technology, where the first *ab initio* Hartree–Fock calculations were carried out using the Roothaan–Hall framework with a Slater-type orbital basis.[16] This led to the birth of computational quantum chemistry as a distinct discipline, which uses electronic structure theory and computers to solve quantum chemical problems.[17]

The 1960s saw a rise in semiempirical methods for quantum chemistry to address the computational challenges of *ab initio* methods. Developed by Roald Hoffmann in 1963, extended Hückel theory (EHT) approximates molecular orbitals by considering only valence electrons. The method populates the diagonal elements of the EHT Hamiltonian with experimental ionization potentials, which serve as proxies for the eigenvalues of the free atom. The off-diagonal elements are then calculated using the Wolfsberg-Helmholz approximation derived from these potentials. Overall this allows expensive integrals to be neglected and replaced by empirical parameters.[18] Different semiempirical methods with different levels of approximation and parameterization strategies followed in the subsequent years, such as complete neglect of differential overlap (CNDO),[19] neglect of diatomic differential overlap (NDDO),[19] intermediate neglect of differential overlap (INDO)[20] and modified intermediate neglect of differential overlap (MINDO).[21] The 1960s and 1970s saw the emergence of density functional theory (DFT), which shifted the focus from wavefunctions to electron density. With the Hohenberg-Kohn theorem in 1964 and the Kohn-Sham equations in 1965, DFT established a theoretical foundation

for calculating ground-state properties based on electron density rather than many-electron wavefunctions.[22,23] DFT provided a computationally efficient alternative to post-Hartree–Fock methods (methods that improve upon Hartree–Fock by accounting for electron correlation), while maintaining reasonable accuracy for many systems. In the following years, the first software packages for quantum chemical calculations were developed. GAUSSIAN70 in 1970 was one of the first general-purpose quantum chemistry programs, making quantum chemical calculations more available to the public.[24]

With the advent of computers, other applications in chemistry emerged. In 1946, King et al. employed IBM business machines to construct rotational spectra of asymmetric rotors, marking the inception of cheminformatics.[25] Cheminformatics is the discipline that applies computer science algorithms to solve chemical problems.[26] During the 1950s and 1960s, chemical data transitioned from printed collections to electronic formats. Zemany introduced the earliest mass spectral search systems using punched cards in 1950,[27] while Kuentzel developed analogous systems for IR spectra in 1951.[28] The 1960s saw notable progress in chemical database retrieval systems, including the establishment of the Chemical Registry System in 1965.[29] Gluck's 1965 algorithm for generating canonical connection tables facilitated compound identification,[30] which Morgan later refined into the widely adopted Morgan Algorithm.[31] Additionally, in 1965, Sussenguth's substructure searching algorithm enhanced efficiency compared to earlier atom-by-atom matching methods.[32]

In cheminformatics, Molecular Design Limited (MDL) Information Systems was founded in 1978 and developed the de-facto standard for representing chemical structures in computer-readable formats. These file formats are part of the chemical table (CT) file family including molecular file (Molfile) and structure-data file (SDfile).[33] Unlike quantum chemistry, cheminformatics uses a representation based on the classical Lewis structure. CT formats depict molecules as undirected graphs with atoms as nodes and bonds as edges. This graphical approach reduces duplicate identification to graph isomorphism, a well-defined mathematical problem. For organic molecules with defined covalent bonds, this method proves particularly effective. Alphanumeric line notations evolved from these graphical representations to provide compact, human-readable molecular identifiers. Instead of storing entire graphs, line notations encode molecular structures as linear strings. The Wiswesser line-formula notation (WLN), introduced in 1949, became the first widely adopted system, encoding structures using digits and letters.[34] Despite its initial success, WLN declined by the 1980s due to its rigid rules and limitations in representing stereochemistry. David Weininger's 1988 introduction of simplified molecular-input line-entry system (SMILES) marked a significant advancement.[35,36] SMILES encodes molecular structures as paths of atomic symbols, with bonds represented implicitly or explicitly, while representing stereocenters and *cis*/*trans* stereochemistry. The SYBYL line notation (SLN) adopted a different approach, explicitly defining aromaticity and avoiding valence assumptions.[37] The representation of structure diagram arranged linearly (ROSDAL), developed for the Beilstein database, provided another specialized notation system.[38] As a response to SMILES' proprietary nature, the international chemical identifier (InChI), devel-

oped since 2000, represents a modern, non-proprietary identifier for chemical substances.[39] Nowadays, SMILES and InChI remain the dominant molecular identifier systems in cheminformatics.

Cheminformatics and quantum chemistry took shape as distinct fields during the 1950s. The former has long depended on the Lewis structure as its foundational molecular representation. The latter, meanwhile, operates within the Born–Oppenheimer approximation and represents molecular structures using three-dimensional (3D) Cartesian coordinates. Cheminformatics typically predicts molecular properties by developing quantitative structure-activity relationship (QSAR) or machine learning (ML) models based on available experimental observations. Quantum chemistry, on the other hand, seeks to resolve chemical problems through approximations of the Schrödinger equation. Since the early 2000s, the divide between these two fields has been steadily narrowing. This development has been driven by the increasing availability of computationally affordable quantum chemical methods, exponential growth in computational power, and by the growing need to analyze and interpret large molecular datasets. Today, standard and user-friendly software packages enable non-experts to routinely perform electronic structure theory calculations on desktop computers. Such calculations are commonly employed for tasks including molecular structure determination, supporting experimental data interpretation, and the prediction of molecular properties and reactivity. A wide range of commercial and open-source software packages supports these applications for molecular systems including `GAUSSIAN`,[24] `ORCA`,[40] `Turbomole`,[41] `TeraChem`[42] and `Q-Chem`[43], among others.

Over the past two decades, semiempirical quantum mechanics (SQM) methods have experienced renewed interest, initiated by the development of the density functional tight binding (DFTB) approach by G. Seifert and coworkers. Despite its success, DFTB relies on extensive atom-pair parameterization, requiring a large number of empirical parameters. This limitation was partially alleviated by the introduction of the GFNn-xTB ($n = 0, 1, 2$) methods by Grimme and coworkers between 2017 and 2019.[44–46] These methods were designed to prioritize the accurate description of molecular geometries, vibrational frequencies, and noncovalent interactions. Unlike earlier semiempirical approaches, they avoid atom-pair parameterization by replacing element-pair parameterized terms with element-wise parameterized EHT and repulsion terms. As a result, a consistent parameterization is achieved across much of the periodic table, extending up to $Z = 86$. In particular, GFN2-xTB has seen widespread adoption since its release in 2019 and is now frequently used as a routine tool for molecular studies. In parallel, automated workflows have been developed to explore chemical space and complex reaction landscapes. Tools such as the global optimization algorithm (GOAT),[47] conformer-rotamer-ensemble sampling tool (CREST)[48] and tensor train conformer search (TTConf)[49] employ semiempirical methods for efficient conformer sampling. Meanwhile, approaches including the *ab initio* nanoreactor method introduced by Martinez and coworkers,[50] Chemoton by M. Reiher and coworkers,[51] the QCxMS2 method by J. Gorges and S. Grimme,[52] and automated reaction search tool (aRST) developed by Y. Chen and C. Bannwarth[53] use quantum-chemical

calculations to autonomously identify and characterize reaction pathways.

The increased accessibility and computational efficiency of electronic structure methods have facilitated the generation of large-scale quantum chemical datasets. Notable among these is PubChemQC,[54] which provides quantum-chemically derived properties for millions of compounds. The QM9 dataset offers DFT-optimized structures and properties for over 134,000 small organic molecules.[55] quantum-mechanical properties of drug-like molecules (QMugs) extends this collection with DFT-optimized structures and properties for more than 665,000 drug-like molecules.[56] For transition metal complexes, the transition metal quantum mechanics (tmQM) dataset includes DFT-optimized structures and properties for over 108,000 such complexes.[57] Recently, the OMol2025 database has expanded this further, containing over 100 million DFT-optimized structures and their associated properties.[58] The availability of these datasets has accelerated the development of cheminformatic ML models that predict molecular properties directly from 3D structures.[59,60] This integration has helped narrow the gap between quantum chemistry and cheminformatics. Quantum chemistry also has increasingly embraced ML techniques to reduce the computational demands of electronic structure calculations. A key application is the development of machine-learning interatomic potential (MLIP) for estimating the electronic energy of molecules without solving the Schrödinger equation directly. Noteworthy examples include neural network potentials by O. Isayev, A. E. Roitberg, and collaborators,[61,62] the SchNet architecture,[63] message passing atomic cluster expansion (MACE)[64] and universal models for atoms (UMA) models,[65] which deliver near-DFT accuracy with significantly lower computational overhead.

This recent progress in theoretical chemistry has produced extensive collections of quantum-chemically derived 3D molecular structures and properties. While early databases integrated experimental data with two-dimensional (2D) CT files, the past two decades have witnessed a rapid increase in quantum-chemically generated 3D structures. In the initial 2D era, structures were manually curated and annotated, whereas today's 3D structures are frequently produced through automated processes. This transition presents a significant challenge: traditional notations such as SMILES and InChI are inherently designed for 2D representations, relying on bond and valence concepts that are not optimized for quantum-chemically derived 3D coordinate data. 3D structures in quantum chemistry data can describe a broad spectrum of molecular structures, including non-standard valence systems like transition states, excited states, and non-covalently bound complexes. Additionally, these structures may represent molecules with inherently complex stereochemistry, such as chiral C_{76} fullerene, where atom-based graph representations are insufficient to capture the full stereochemical complexity. Despite this diversity, all structures must be uniquely identifiable, searchable in databases, and representable in a compact format. As quantum chemistry continues to populate entire databases, the demand for identifier systems capable of directly handling 3D coordinates has become urgent.

2 Theory

2.1 Quantum Chemistry and Potential Energy Surface

In non-relativistic quantum chemistry, systems described by a set of coordinates $\mathbf{r}$ are characterized by complex-valued wave functions $\Psi(\mathbf{r}, t)$. The wave function $\Psi(\mathbf{r}, t)$ encodes all information about the system. Since the non-relativistic Hamilton operator is independent of electron spin, pure spin eigenfunctions are assumed for the wave function. Consequently, the total wave function can be expressed as a product of a spatial part and a spin part. In the following, only the spatial part depending on the coordinates $\mathbf{r}$ is considered, as it contains all the relevant parameters for the non-relativistic Hamiltonian. While the wave function itself has no direct physical interpretation, its modulus squared $|\Psi(\mathbf{r}, t)|^2$ provides the probability density of finding particles at specific positions and times. Normalization requires that the integral of the probability density over all space equals one:

$$\int |\Psi(\mathbf{r}, t)|^2 \, \mathrm{d}\mathbf{r} = 1. \tag{2.1}$$

These wave functions are solutions to the Schrödinger equation and evolve over time t. The time-dependent Schrödinger equation is expressed as:[66]

$$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{r}, t) = \hat{H} \Psi(\mathbf{r}, t). \tag{2.2}$$

$\hat{H}$ represents the Hamilton operator, which describes the total energy of the system. This operator is Hermitian and comprises kinetic and potential energy terms for all particles in the system:

$$\hat{H} = \hat{T} + \hat{V}. \tag{2.3}$$

In most chemical systems of interest, the potential energy operator is time-independent. For such operators, the time-dependent Schrödinger equation can be separated into spatial and temporal components. By applying this *ansatz* for the wave function:

$$\Psi(\mathbf{r}, t) = \psi(\mathbf{r}) \cdot \phi(t), \tag{2.4}$$

the time-dependent Schrödinger equation reduces to the time-independent Schrödinger equation for the spatial part:

$$\hat{H} \psi(\mathbf{r}) = E \psi(\mathbf{r}). \tag{2.5}$$

E represents the energy eigenvalues of the system. The time-dependent component simplifies to a phase factor:

$$\phi(t) = e^{-\frac{iEt}{\hbar}}, \tag{2.6}$$

which is typically omitted when solving time-independent problems.[66] In such cases, the Hamilton operator $\hat{H}$ can be expressed as the sum of kinetic and potential energy operators:

$$\hat{H} = \hat{T}_{\text{elec}} + \hat{T}_{\text{nuc}} + \hat{V}_{\text{elec-elec}} + \hat{V}_{\text{nuc-nuc}} + \hat{V}_{\text{elec-nuc}}. \tag{2.7}$$

$\hat{T}_{\text{elec}}$ and $\hat{T}_{\text{nuc}}$ denote the kinetic energy operators for electrons and nuclei, respectively. $\hat{V}_{\text{elec-elec}}$, $\hat{V}_{\text{nuc-nuc}}$ and $\hat{V}_{\text{elec-nuc}}$ correspond to the potential energy operators for electron-electron, nucleus-nucleus, and electron-nucleus interactions, respectively. The general form of the kinetic operator is:

$$\hat{T} = -\frac{\hbar^2}{2m}\nabla^2, \tag{2.8}$$

where m is the mass of the particle (electron or nucleus), and ∇^2 is the Laplace operator representing the second spatial derivatives. Potential energy operators are typically defined by Coulomb interactions between charged particles:

$$\hat{V}_{ij} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \tag{2.9}$$

Here, q_i and q_j are the charges of particles i and j, ϵ_0 is the permittivity of free space, and r_{ij} is the distance between the particles.

In computational chemistry, the Born–Oppenheimer approximation is frequently applied due to the significant mass difference between electrons and nuclei. This approximation assumes that nuclei remain stationary relative to the rapidly moving electrons. Electrons are thus treated as moving in the field of fixed nuclei. The electronic and nuclear Schrödinger equations can therefore be decoupled. The total wave function is expressed as a product of electronic and nuclear wave functions:

$$\Psi(\mathbf{r}, \mathbf{r}_{\text{nuc}}) = \psi_{\text{elec}}(\mathbf{r}; \mathbf{r}_{\text{nuc}}) \cdot \chi_{\text{nuc}}(\mathbf{r}_{\text{nuc}}). \tag{2.10}$$

$\psi_{\text{elec}}(\mathbf{r}; \mathbf{r}_{\text{nuc}})$ is the electronic wave function, dependent on electronic coordinates $\mathbf{r}$ and parametrically on nuclear coordinates $\mathbf{r}_{\text{nuc}}$. Here, χ_{nuc} is the nuclear wave function, dependent solely on nuclear coordinates. Substituting this *ansatz* into the time-independent Schrödinger equation yields two distinct equations: one for the electronic part and one for the nuclear part. The electronic Schrödinger equation is:

$$\hat{H}_{\text{elec}}\psi_{\text{elec}}(\mathbf{r}; \mathbf{r}_{\text{nuc}}) = E_{\text{elec}}\psi_{\text{elec}}(\mathbf{r}; \mathbf{r}_{\text{nuc}}). \tag{2.11}$$

$\hat{H}_{\text{elec}}$ is the electronic Hamilton operator, equivalent to Equation 2.7 neglecting the nuclear kinetic energy operator $\hat{T}_{\text{nuc}}$. E_{elec} denotes the electronic energy. Here, $\hat{V}_{\text{nuc-nuc}}$ becomes a

constant, affecting the electronic energy but not the electronic wave function. The nuclear Schrödinger equation is:

$$\left(\hat{T}_{\mathrm{nuc}} + E_{\mathrm{elec}}\right) \chi_{\mathrm{nuc}}(\mathbf{r}_{\mathrm{nuc}}) = E\chi_{\mathrm{nuc}}(\mathbf{r}_{\mathrm{nuc}}). \tag{2.12}$$

The Born–Oppenheimer approximation neglects coupling between electronic and nuclear motions, greatly simplifying calculations, while introducing only negligible errors for energetically well-separated electronic states. For a given set of nuclear coordinates, the electronic Schrödinger equation can thus be solved to determine the electronic energy E_{elec}.[66]

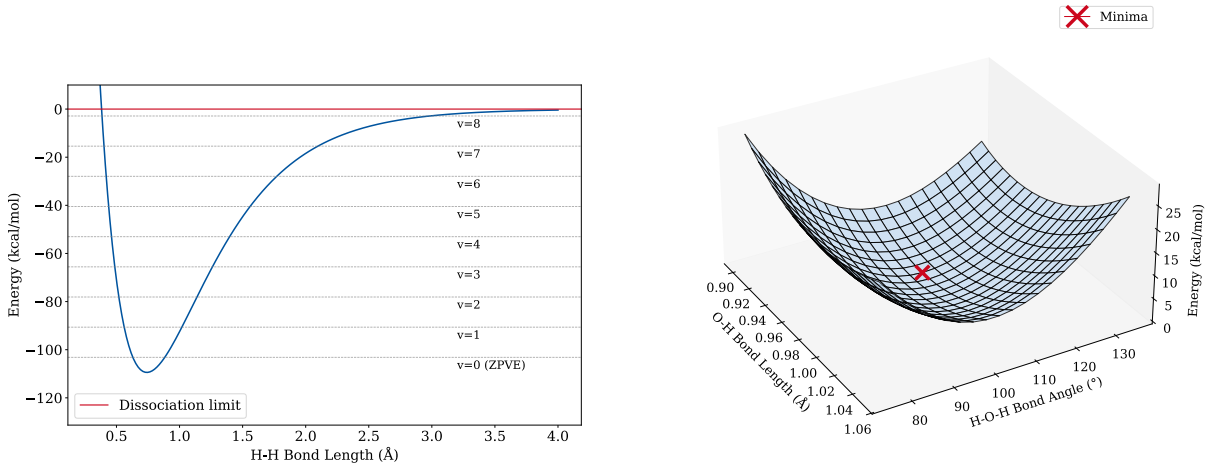
The electronic Schrödinger equation presents a many-body problem for which exact solutions are intractable when more than one electron is involved. A variety of methods have been developed to address this challenge, with the Hartree–Fock method serving as a fundamental starting point. Hartree–Fock employs a mean-field approximation, treating each electron as moving independently in an average field created by all other electrons. This approach neglects instantaneous electron-electron interactions, known as electron correlation effects, but simplifies the many-electron problem to a computationally manageable form. While Hartree–Fock captures approximately 99% of the total energy, the exclusion of electron correlation can lead to significant errors in relative energy and property calculations. Post-Hartree–Fock methods, such as configuration interaction (CI)[67] and coupled cluster (CC),[67] refine this framework by explicitly incorporating electron correlation. Among these, coupled cluster methods like CCSD(T) are regarded as the "gold standard" in quantum chemistry due to their high accuracy in predicting molecular properties. However, these methods demand substantial computational resources, often restricting their use to smaller molecular systems. As an alternative, density functional theory (DFT) offers a computationally efficient solution by reformulating the many-electron problem in terms of electron density rather than the wave function.[68] Electron correlation in DFT is addressed through exchange-correlation functionals, which approximate complex electron interactions. The accuracy of DFT calculations depends heavily on the choice of the functional. This has led to the development of numerous functionals tailored to specific systems and properties. With its favorable computational scaling, DFT enables the study of larger molecular systems compared to post-Hartree–Fock approaches. It has thus become a standard tool for mechanistic studies in both organic and inorganic computational chemistry. For high-throughput screening or larger systems, SQM methods such as GFN2-xTB[46] or g-xTB[69] further reduce computational demands by introducing empirical parameters to replace expensive integrals. When selecting a method from this range of techniques, the specific chemical problem, the desired accuracy, and the available computing resources must be taken into account to strike a balance between accuracy and computational feasibility. While this thesis does not delve into method development or their direct application, these techniques are implemented in widely used quantum chemistry software, including `Gaussian`,[24] `ORCA`,[40] `Turbomole`,[41] and `xTB`.[70] These packages enable streamlined, nearly black-box calculations, requiring minimal input beyond the molecular structure. As a result, theoretical calculations

have become an integral part of chemical research as standard tools for interpreting experimental data, predicting molecular properties, and elucidating reaction mechanisms.

As these methods approximately solve the electronic Schrödinger equation for fixed nuclear positions, they yield the electronic energy E_{elec} as a function of nuclear coordinates. These energies define the potential energy surface (PES) of a molecular system. The PES serves as a core concept in quantum chemistry, illustrating how a molecular system's energy varies with nuclear positions. It forms a multidimensional surface where each point reflects a specific nuclear arrangement, and the value at that point indicates the system's electronic energy. Therefore, it is the central concept for understanding chemical reaction pathways and provides estimates for the thermodynamics and kinetics of reactions.

Instead of using Cartesian coordinates to describe nuclear positions, internal coordinates provide a more intuitive description of the PES. Internal coordinates q generally encompass bond lengths, bond angles, and dihedral angles or combinations thereof. For a molecule with n_{atoms} atoms, $3n_{\mathrm{atoms}} - 6$ internal coordinates describe non-linear molecules, while $3n_{\mathrm{atoms}} - 5$ are needed for linear molecules. Internal coordinates inherently exclude translational and rotational degrees of freedom, focusing exclusively on structural changes. For the dihydrogen molecule (H_2), only one internal coordinate is needed instead of six Cartesian coordinates. The PES simplifies to a one-dimensional curve, as only the bond length $r_{\mathrm{H\text{-}H}}$ is required to describe the system. Figure 2.1a displays a schematic potential energy curve of the H_2 molecule as a function of its bond length. Elongating a bond beyond its equilibrium length increases the system's energy until bond dissociation occurs, reaching the dissociation limit. At this limit, the energy corresponds to that of two non-interacting hydrogen atoms. The bond dissociation energy is defined as the energy difference between the equilibrium bond length and this dissociation limit. Compressing the bond length also raises the energy due to repulsive nuclear forces. The equilibrium bond length coincides with the minimum of the PES, where nuclear forces are balanced and the gradient $\mathrm{d}E_{\mathrm{elec}}/\mathrm{d}q$ equals zero. The curve in Figure 2.1a represents the electronic energy derived from solving the electronic Schrödinger equation. Solving the nuclear Schrödinger equation on this PES yields the quantized nuclear energy levels. In the harmonic approximation shown in Figure 2.1a, these correspond to equally spaced vibrational levels with quantum numbers $v = 0, 1, 2, ...$ Due to the Heisenberg Uncertainty Principle, the nuclei cannot be localized at the potential minimum with zero momentum; consequently, even at 0 K the molecule retains a zero-point energy (ZPE). This represents the $v = 0$ vibrational level, which is the lowest energy state of the system. It consists of the electronic energy at the minimum plus the zero-point vibrational energy with both kinetic and potential energy contributions from the vibrational motion of the nuclei.[71]

The PES for water (H_2O) increases in complexity due to its dependence on three internal coordinates. In the following, the PES is represented as a function of two O-H bond lengths and the H-O-H bond angle. Figure 2.1b depicts the two-dimensional PES of water, illustrating energy variations with changes in bond lengths and bond angle. For clarity, this representation

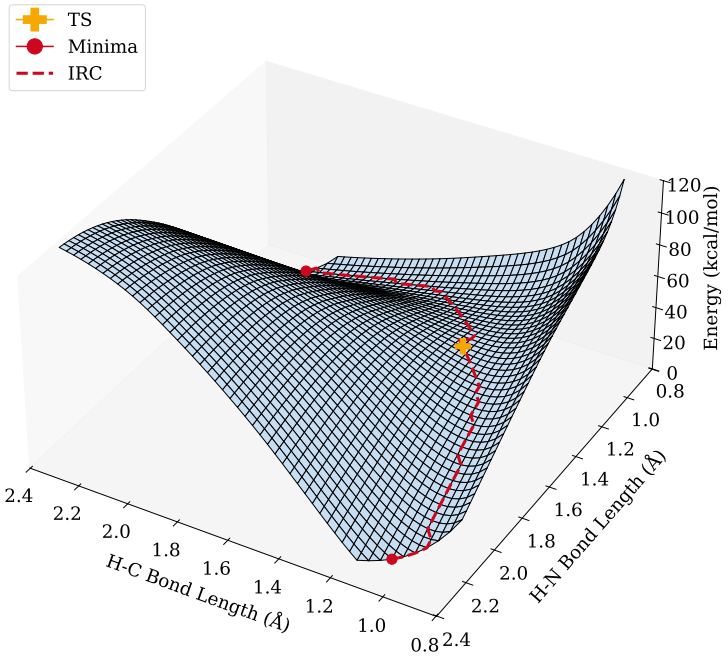


(a) Potential Energy Curve for H_2

(b) Potential Energy Surface for H_2O

Figure 2.1: Schematic PES for H_2 and H_2O: (a) The quantitative one-dimensional PES of H_2 as a function of the H-H bond length. The vibrational energy levels are calculated using the harmonic approximation at the minimum structure. These levels are indicated as dashed lines, with the lowest level representing the zero-point energy (ZPE). (b) The two-dimensional PES of H_2O as a function of the O-H bond length and the H-O-H bond angle. The energy is calculated using the MLIP UMA-S[65] on a grid of internal coordinates. The minimum of the surface corresponds to the equilibrium geometry of water.

assumes equal O-H bond lengths, reducing the coordinates to one O-H bond length and the H-O-H bond angle. The PES is approximately calculated using the MLIP UMA-S[65] on a grid of internal coordinates without geometry relaxation. The PES of water features a minimum that aligns with its equilibrium geometry, where O-H bond lengths and the H-O-H bond angle reach their optimal values. This minimum represents the most stable configuration, where nuclear forces are balanced. Any deviation from this geometry results in higher energy, indicating less stable configurations. The minimum is classified as a stationary point on the PES. Stationary points are defined as configurations where the gradient of the PES with respect to all internal coordinates equals zero. Such stationary points are categorized as minima, transition states, or saddle points according to the curvature of the PES at those locations. The curvature is determined by the second-order derivatives of the energy with respect to the internal coordinates, forming the Hessian matrix **H**. Minima represent stable configurations where all curvatures are positive (i.e., all eigenvalues of the Hessian matrix are positive), indicating that energy increases in every direction away from the minimum. First-order saddle points or transition states act as energy barriers between minima and exhibit one negative curvature direction (i.e., one negative eigenvalue of the Hessian matrix), corresponding to the reaction coordinate along which energy decreases toward reactants and products. Second-order saddle points possess two negative curvature directions (i.e., two negative eigenvalues of the Hessian matrix) and are generally not directly involved in chemical reactions. In chemical reactions, minima and transition states are especially relevant, as they define stable states and the energy barriers that must be surmounted during the reaction. The energy difference between a minimum and a transition state along the reaction coordinate determines the activation energy of



(a) Potential Energy Surface for HCN to HNC Isomerization



(b) Minimum HCN (front)



(c) Transition State



(d) Minimum HNC (back)

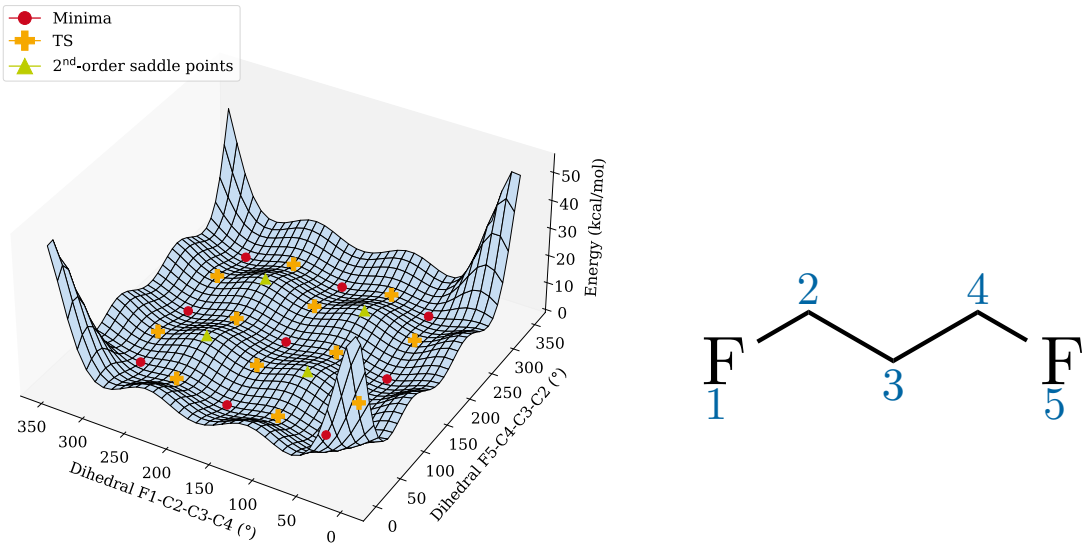
Figure 2.2: Approximated PES for the isomerization of HCN to HNC was calculated using the MLIP UMA-S[65] on a grid of H-C and H-N bond lengths. The C-N bond length was kept fixed at 1.149 Å. Energy is given in kcal/mol relative to the minimum energy of HCN. The minimum in front corresponds to the equilibrium geometry of HCN (b). The minimum at the back represents the equilibrium geometry of HNC (d). The first-order saddle point in the middle represents the transition state for the isomerization (c), characterized by one negative eigenvalue of the Hessian matrix, indicating a maximum along the reaction coordinate and minima in all other directions. The intrinsic reaction coordinate (IRC) (dashed line) is also approximated on the grid and connects the two minima via the transition state.

the reaction. This activation energy governs the reaction kinetics, influencing the reaction rate. Lower activation energies result in faster reactions, while higher activation energies slow the reaction. The energy difference between two minima establishes the reaction's thermodynamics, revealing whether it is exothermic or endothermic.[72]

Figure 2.2 presents a simplified PES for the isomerization of Hydrogen Cyanide (HCN) to Hydrogen Isocyanide (HNC) via a 1,2-hydrogen shift. The PES is calculated using the MLIP UMA-S.[65] It is depicted as a two-dimensional surface based on two internal coordinates. The selected internal coordinates for this representation are the H-C and H-N bond lengths, with the C-N bond length held fixed. Individual points on the PES are derived through a grid-based method, where each combination of H-C and H-N bond lengths defines a geometry for which the energy is computed using the MLIP. The PES features two minima, corresponding to the equilibrium geometries of HCN and HNC. HCN, being the more stable isomer, represents the

global minimum on the PES due to its lower electronic energy. HNC, less stable and with higher electronic energy, corresponds to a local minimum. The first-order saddle point between these minima indicates the transition state for the isomerization process. A dashed line marks the intrinsic reaction coordinate (IRC), which connects the two minima on the potential energy surface via the transition state. The minimum energy reaction pathway (MERP) is defined as the steepest-descent path of the potential energy surface, starting from the transition state and following the direction of the eigenvector corresponding to the negative eigenvalue of the Hessian matrix. The IRC corresponds to the MERP expressed in mass-weighted coordinates, thereby linking the transition state to the reactant and product minima in a physically meaningful manner. This path can be interpreted as the trajectory followed by a molecular system moving downhill from the transition state along the reactant and product valleys with initial zero kinetic energy.[73] It signifies the most probable reaction pathway on the PES. The isomerization of HCN to HNC serves as an example of a chemical reaction on a PES. Chemical reactions involve the breaking and formation of bonds, corresponding to transitions across the PES. Reactants and products are represented by distinct minima on the PES. However, these minima do not always represent different chemical species but can also indicate various conformations of the same molecule. Conformations are spatial arrangements that interconvert via rotations around single bonds.

An example of conformational changes on a PES is provided by Figure 2.3. It illustrates conformations of a propane derivative with two fluorine substituents, plotting the PES as a function of the dihedral angles F1-C2-C3-C4 and F5-C4-C3-C2. The PES features multiple minima, each corresponding to a different conformer of the molecule. Interconversion between conformers occurs through rotation around the C2-C3 and C3-C4 single bonds. Conformational minima are typically separated by low energy barriers, representing the rotational energy required for interconversion. For the propane derivative in Figure 2.3, these barriers are around 3 kcal/mol. Such low barriers allow rapid interconversion at room temperature, occurring on nanosecond to microsecond timescales. In contrast, the isomerization of HCN to HNC involves a significantly higher barrier of about 48 kcal/mol, as depicted in the approximated PES in Figure 2.2. This high barrier effectively prevents spontaneous isomerization at room temperature. The boundary between conformers and isomers is continuous rather than sharply defined. The activation barrier for dihedral rotation can vary widely, ranging from a few kcal/mol to over 30 kcal/mol. When these barriers are sufficiently high, conformers can be isolated as distinct chemical species, behaving as separate molecules. Ōki proposed a practical guideline in the context of atropisomerism: distinct isomers exhibit a half-life of at least 1,000 seconds for rotational conversion under ambient conditions.[74] This corresponds to an energy barrier of roughly 22 kcal/mol to separate conformers from configurational isomers.



(a) PES **(b)** Lewis structure of the propane derivative

Figure 2.3: PES for a propane derivative with two fluorine substituents as a function of the dihedral angles F1-C2-C3-C4 and F5-C4-C3-C2. The PES was calculated using the MLIP UMA-S[65] on a grid of dihedral angles. Energy is given in kcal/mol relative to the minimum energy conformation. Multiple minima correspond to different conformers of the molecule, which can interconvert by rotation around the C2-C3 and C3-C4 single bonds. The energy barriers between these conformers are relatively low, allowing for rapid interconversion at room temperature. Minima are stable conformers with no negative eigenvalues of the Hessian. First-order saddle points/transition states exhibit one negative eigenvalue of the Hessian, while second-order saddle points have two negative eigenvalues of the Hessian. Transition states and second-order saddle points on the periodic boundaries are not shown for clarity.

2.2 Molecular Representations in Quantum Chemistry and Cheminformatics

Molecular structures generated by quantum chemical or cheminformatics workflows must be stored in formats compatible with diverse software tools for visualization, analysis, and further processing. Numerous file formats have emerged in both disciplines since the 1970s, but none has become a universal standard for both.[33] This fragmentation stems from divergent requirements in both fields. According to the Born–Oppenheimer approximation, the electronic Schrödinger equation requires only the 3D atomic positions (in addition to the method, the basis set, the total charge, and the multiplicity). Thus, quantum chemistry prioritizes 3D atomic coordinates, which are stored in respective file formats. No other explicit structural information is necessary for quantum chemistry. On the other hand, cheminformatics requires connectivity and stereochemical information in the file formats, as many algorithms depend on graph representations of molecules. Cheminformatics does not necessarily require 3D coordinates; instead, it often only uses 2D representations of Lewis structure depictions. While quantum chemistry implicitly represents connectivity and stereoinformation in the form of 3D coordinates, cheminformatics formats must explicitly encode this information. Other fields, such as crystallography, have their own distinct requirements. These fundamental differences in data representation mean that a unified standard has yet to be established.[33]

In quantum chemistry, 3D coordinates are typically stored in simple text-based file formats. In the following, two common formats for storing 3D coordinates in quantum chemistry are described. The XYZ format and the `TURBOMOLE` coord file format are shown in Figure 2.4b and Figure 2.4c, respectively. To explain the structure of these file formats, numbered annotations are used in the figures and in the following text. In the XYZ format, the first line indicates the number of atoms (1), followed by an optional comment line (2). Subsequent lines list each atom's element symbol (3) and its Cartesian coordinates in Å (4). The `TURBOMOLE` coord file format[75] starts with the keyword "$coord", followed by lines containing the Cartesian coordinates in Bohr units (5) and the corresponding element symbols (6). Such file formats are straightforward and widely accepted by various quantum chemistry software packages such as `ORCA`,[40] `TURBOMOLE`,[41] and `xTB`.[70] Another relevant file format in quantum chemistry is the Z-matrix format, which represents the spatial arrangement of atoms using internal coordinates, including bond lengths, bond angles, and dihedral angles. For each new atom, the Z-matrix first specifies its distance to a previously defined atom in the matrix, followed by the bond angle relative to a second atom, and finally the dihedral angle relative to a third atom. Notably, the specified distance does not necessarily imply a chemical bond between the atoms. It also allows for a compact representation to build 3D structures by manual input. Other file formats, specific to certain quantum chemistry software packages, may be used to include not only 3D coordinates but also additional information for carrying out the electronic structure calculations, such as basis sets, charge, multiplicity, and other calculation parameters.[76]

In cheminformatics, the most widely used file format is the Molfile originally developed by MDL. After several acquisitions, MDL is now part of BIOVIA, a subsidiary of *Dassault Systèmes*. The Molfile was first publicly described in 1992 but developed in the years prior.[77] The Molfile exists in two versions: V2000 and V3000 (see Figure 2.4d and Figure 2.4e, respectively). Both versions start with a header section. The first line can contain the molecule name but is often left blank. The second line follows a strict format: user initials, the program name, date of creation, and whether the file is a 2D or 3D representation (7 or 18, respectively). The third line is usually empty or contains comments.[78] After the header section, the connectivity table (Ctab) section begins. The first important line of the Ctab section is the counts line. It contains information about the number of atoms and bonds in the molecule. In V2000, the line starts with the number of atoms (8) and the number of bonds (9). A chiral flag (10) indicates whether the molecule contains stereochemical information. The line ends with the version of the Molfile format (11). In V3000, the counts line starts with "M V30 COUNTS" followed by the number of atoms (20) and bonds (21) and a chiral flag (22). Here, the file format version has been moved to the header section (19). Other information can also be found in both counts lines, but it is not discussed here in detail, as most of it is rarely used.[78]

In both versions, the atom block follows next. Each atom is specified by its coordinates (12) in Ångströms and its element symbol (13). Not all files provide 3D coordinates; frequently, the z-coordinate is set to zero, as they are generated by 2D drawing programs. These coordinates reflect the 2D arrangement of the molecule in these Lewis structure depictions, which may not correspond to the actual 3D geometry of the molecule. For V3000, each line in the atom block begins with "M V30", followed by the atom index, element symbol (23), and coordinates (24).[78]

A notable difference is that both Molfile formats in Figure 2.4d/2.4e usually omit hydrogen atoms, unlike the XYZ and `TURBOMOLE` coord file formats where hydrogen atoms are always included. Hydrogen atoms are usually excluded from CT files for clarity, as they can be inferred from valency rules. In both V2000 and V3000, the atom block includes additional columns for isotopes, charges, and radical states. In V2000, one such column is the stereo flag (14), which classifies an atom as achiral, odd, even, or either (0, 1, 2, 3, respectively). For (2R,3E)-3-Penten-2-ol, the carbon atom at position 4 is chiral and carries an even stereo flag.[78] In V3000, this column is replaced by the CFG keyword (25), which encodes the same stereochemical information as the stereo flag in V2000.[78]

Following the atom block, the bond block appears in both formats. Each line specifies the two bonded atoms and the bond type. In V2000, the first two columns list the indices of the bonded atoms (15). The third column (16) defines the bond type (single, double, triple, aromatic, single/double, single/aromatic, double/aromatic, any) using integers 1–8. Stereochemistry of bonds is also encoded in the bond block. The fourth column (17) is the bond stereo indicator, where the numbers 0, 1, 4, and 6 denote non-stereo, up, either, and down configurations, respectively. For double bonds, 0 means to use the Cartesian coordinates from the atom block

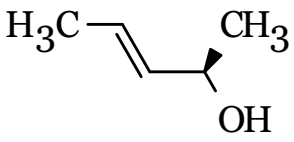
to determine the E/Z configuration, while 3 indicates that the double bond is either *E* or *Z*. In V3000, the bond block begins with the bond index, followed by the bond type (26) and the indices of the bonded atoms (27). Unlike V2000, V3000 supports bond properties via keyword-value pairs after the atom indices. For example, CFG denotes bond configuration (none, up, either, down) as 0–3 (28).[78]

V2000 was developed when punched cards and the programming language `Fortran` were widely used. Hence, these early file formats are quite restrictive, e.g. each line cannot exceed 80 characters. Further, many fields have fixed widths, making the position of blank spaces important. The V3000 format addresses the constraints of V2000. V2000's fixed-width fields impose significant restrictions. For instance, the fields for atom and bond counts are limited to three characters, capping the maximum number of atoms and bonds at 999. Coordinate fields are restricted to five digits before the decimal (including the sign) and four digits after. V3000 resolves these issues by adopting variable-width fields. It also introduces two additional bond types: coordination and hydrogen bonds. The format supports haptic bonds, where multiple atoms in a π-system coordinate to a single atom, as seen in ferrocene. Subgroups within molecules can be defined to highlight substructures in visualization tools or to enhance stereochemical specifications. Enhanced stereochemistry enables the representation of stereoisomer mixtures, for example racemic mixtures.[78]

Several file formats have been developed in the 1980s by MDL for representing chemical structures. All these file formats are grouped under the umbrella of chemical table (CT) files, where the core is the Ctab. One of them is the structure-data file (SDfile) as a Molfile extension. The Molfile itself only represents a single molecule. In contrast, the SDfile allows storing multiple molecules in one file by concatenating multiple Molfile entries. Further, it allows adding additional data fields for each molecule, such as physical properties, identifiers, or annotations. It is especially valuable for managing large molecular datasets with supplementary properties or annotations. Related formats like RGFile (for queries) and RXNFile (for reactions) exist but are not covered here.[78] Other notable file formats developed by different organizations such as Protein Data Bank (PDB)[79] for proteins, crystallographic information file (CIF) for crystallographic data,[80] and chemical markup language (CML),[81] an XML-based format for chemical data, are also widely used in cheminformatics but are not discussed in detail here either.

All these file formats have in common that they are rather large in size, as they store full coordinates and connectivity information. For many applications in cheminformatics, more compact representations are preferred. Such representations are molecular string notations that encode the molecular structure in a single line of text. The most prominent examples are SMILES and InChI.

David Weininger created the SMILES notation in 1988 to provide a simple and human-readable way to represent chemical structures as strings.[35] It is a highly compact form of describing molecular structures as graphs. The SMILES code has been widely adopted in cheminformatics as an exchange format for chemical structure and ML applications due to its simplicity



(a) (2R,3E)-3-Penten-2-ol molecule

```
16[1]
(2R,3E)-3-Penten-2-ol[2]
C[3]        6.57349[4]        2.64193        0.34225
C           6.07299           1.24659        0.15732
C           5.27666           0.62483        1.04336
C           4.78065          -0.78101        0.87835
O           3.35906          -0.75422        0.85026
C           5.24485          -1.67666        2.01926
H           7.24345           2.90846       -0.48140
H           7.13964           2.73343        1.27446
H           5.74567           3.35722        0.34682
H           6.38018           0.73267       -0.75001
H           4.96985           1.15483        1.94197
H           5.11601          -1.20835       -0.07336
H           3.05257          -0.42514        1.71104
H           4.87749          -1.31297        2.98624
H           6.33711          -1.73547        2.06223
H           4.84501          -2.68885        1.89427
```

(b) XYZ file format. The first line (1) indicates the number of atoms, followed by a comment line (2). Each subsequent line lists the element symbol (3) and Cartesian coordinates in Å (4).

```
$coord
   12.22377493492663     -3.78268614753223     -5.38998128785289[5]     c[6]
   11.45198406531726     -1.19058892874372     -4.58553749078401     c
   10.02647524450735     -0.73334439847436     -2.53539633470001     c
    9.15787558535779      1.79568682116667     -1.57210364164250     c
    6.46725627321713      1.73621900737307     -1.51244731687171     o
    9.98522136305661      4.09894873271537     -3.09119738179569     c
   13.47570553078321     -3.66075217335877     -7.03296625965879     h
   10.57474840865419     -4.90596148813913     -5.93487749552734     h
   13.25514620783336     -4.75455954219643     -3.88383590248232     h
   12.11822988501096      0.35390520359679     -5.76819959846551     h
    9.40463269570594     -2.33118182100367     -1.39418644004717     h
    9.81068666144376      2.01400274240458      0.38079403817233     h
    5.90535461460354      1.42857265428178     -3.23562157682643     h
    9.29799738833911      4.00147709663490     -5.04275356886613     h
   12.04445892864137      4.29079396071717     -3.11718260047344     h
    9.17823491141324      5.82323002210436     -2.27648206848834     h
$end
```

(c) TURBOMOLE coord file format. After the keyword "$coord", each line contains Cartesian coordinates in Bohr units (5) followed by the element symbol (6).

```
OpenBabel01072609212D[7]

 6[8]  5[9]  0  0  1[10]  0  0  0  0  0999 V2000[11]
   2.5981   -1.5000    0.0000[12] C[13]  0  0  0   0  0  0  0  0  0  0  0  0
   1.7321   -1.0000    0.0000   C    0  0  0   0  0  0  0  0  0  0  0  0
   0.8660   -1.5000    0.0000   C    0  0  0   0  0  0  0  0  0  0  0  0
  -0.0000   -1.0000    0.0000   C    0  0  2[14]  0  0  0  0  0  0  0  0  0
  -0.8660   -1.5000    0.0000   O    0  0  0   0  0  0  0  0  0  0  0  0
  -0.0000   -0.0000    0.0000   C    0  0  0   0  0  0  0  0  0  0  0  0
 1  2[15]  1   0  0  0  0
 2  3   2[16]  0  0  0  0
 3  4   1   0  0  0  0
 4  5   1   0  0  0  0
 4  6   1   6[17]  0  0  0
M  END
```

(d) MDL Molfile format V2000. After the header (7), the counts line specifies the number of atoms (8), bonds (9), chiral flag (10), and format version (11). The atom block lists coordinates (12) and element symbols (13). Stereo flags (14) indicate chirality. The bond block defines bonded atoms (15) and bond types (16). Bond stereochemistry is encoded in the bond block (17) using indicators for non-stereo, up, either, and down configurations.

```
OpenBabel01072609302D[18]

  0  0  0     0  0            999 V3000[19]
M  V30 BEGIN CTAB
M  V30 COUNTS 6[20] 5[21] 0 0 1[22]
M  V30 BEGIN ATOM
M  V30 1 C[23] 2.59808 -1.5 0 0[24]
M  V30 2 C   1.73205 -1.0 0 0
M  V30 3 C   0.86602 -1.5 0 0
M  V30 4 C   0.00000 -1.0 0 0 CFG=2[25]
M  V30 5 O   -0.8660 -1.5 0 0
M  V30 6 C   -0.0000 -0.0 0 0
M  V30 END ATOM
M  V30 BEGIN BOND
M  V30 1 1   1 2
M  V30 2 2[26] 2 3[27]
M  V30 3 1   3 4
M  V30 4 1   4 5
M  V30 5 1   4 6 CFG=3[28]
M  V30 END BOND
M  V30 END CTAB
M  END
```

(e) MDL Molfile format V3000. The header includes the format version (19). The counts line specifies atom count (20), bond count (21), and chiral flag (22). The atom block lists element symbols (23) and coordinates (24). The chirality flag for the stereocenter in the atom block is encoded by the CFG keyword (25). The bond block defines bond types (26) and bonded atoms (27). Bond properties such as configuration can be specified via keyword-value pairs, e.g. CFG=3 (28) indicates a "down" configuration for the bond between atoms 4 and 6.

Figure 2.4: Common file formats used in quantum chemistry and cheminformatics, illustrated with (2R,3E)-3-Penten-2-ol: (a) molecular structure; (b) XYZ format with Cartesian coordinates in Å; (c) TURBOMOLE coord format with coordinates in Bohr units; (d) MDL Molfile V2000 format; (e) MDL Molfile V3000 format. The Molfile formats encode connectivity, bond types, and stereochemical information explicitly, while coordinate-based formats represent this information implicitly through atomic positions.

and ease of use. The input to generate the SMILES is typically the discussed CT file format as SMILES does not come with algorithms to generate the connectivity and stereoinformation from 3D coordinates, but rather relies on predefined connectivity and stereoinformation. SMILES relies on a few basic rules to convert a molecular structure into a string representation. Atoms are represented by their atomic symbols, e.g. C for carbon, O for oxygen, N for nitrogen. Hydrogen atoms are usually not included explicitly (though they can be), but are inferred based on the valency of the heavy atoms. Charges are indicated by "+" or "-" symbols following the atomic symbol. To represent a molecule, a path through the molecular graph is followed. This means that neighboring atoms are written one after another. Branches in the molecular structure are indicated by parentheses. Rings are represented by numbers that indicate the start and end of the ring (e.g. C1CCCCC1 for cyclohexane). Only double bonds (=), triple bonds (#) are explicitly indicated, while single and aromatic bonds are usually omitted. Aromatic atoms are represented by lowercase letters (e.g. c for aromatic carbon). Separate fragments are divided by a dot (.). Stereochemistry is encoded by further symbols. The configuration of stereocenters is indicated by the "@" symbol. When reading the SMILES string from left to right, it describes the arrangement of the atoms after the "@" symbol. Here, "@" indicates a counter-clockwise arrangement, while "@@" indicates a clockwise arrangement. Double bond stereochemistry is indicated by "/" and "\" symbols to denote the relative positions of substituents. A few other forms such as allene, square-planar, trigonal-bipyramidal, and octahedral stereochemistry can be encoded by some implementations such as the one from Daylight.[82]

A given molecular structure often corresponds to a large number of valid SMILES representations. For a molecule comprising n_{atoms}, up to $n_{\mathrm{atoms}}!$ distinct Ctab representations may exist, excluding symmetry considerations. Thus for SMILES, traversing the molecular graph from different starting points or along different paths can yield multiple valid SMILES strings for the same molecule. Canonicalization algorithms select one specific SMILES from all possible valid representations, designating it as the "canonical SMILES". The primary objective of canonicalization is to assign a unique atom numbering to a molecular structure. This ensures a single, consistent SMILES representation for that structure, regardless of the original input format. David, Arthur, and Joseph Weininger introduced one such algorithm in the original Daylight implementation of SMILES.[36] Another notable canonicalization method is the Morgan algorithm, developed by H. L. Morgan in 1965.[31] However, different software packages employ proprietary canonicalization algorithms, potentially producing varying canonical SMILES for identical molecules depending on the implementation. Consequently, canonical SMILES are specific to the software used for their generation. Next to canonical SMILES, several other variants exist. SMILES that encode only the labeled molecular graph but omitting chiral or isotopic details are referred to as "generic SMILES".[82] When isotopic and chiral specifications are included, the resulting SMILES are collectively termed "isomeric SMILES". If aromaticity is depicted using alternating single and double bonds instead of lowercase aromatic symbols, the representation is known as "Kekulé SMILES". Table 2.1 shows several examples of SMILES

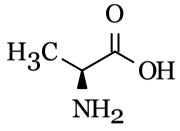
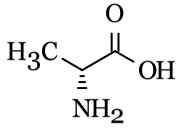
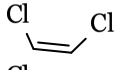
representations for different molecules, demonstrating the SMILES grammar for features such as implicit hydrogens, double bonds, triple bonds, rings, aromaticity, and stereochemistry.

During the early 2000s, the chemistry community faced a substantial issue related to interoperability. Chemical databases depended on proprietary systems, including CAS Registry Numbers and various SMILES implementations. Additionally, the majority of methods at the time characterized molecules based on their input Lewis structure representation. This approach resulted in mesomeric forms or tautomers being assigned distinct identifiers. Such a practice proved problematic for database searching and indexing, given the equivalence and dynamic interconversion of these forms. To address these issues, the International Union of Pure and Applied Chemistry (IUPAC) and the National Institute of Standards and Technology (NIST) introduced the IUPAC international chemical identifier (InChI) as a non-proprietary, open-source standard for chemical structure representation. Originally known as the IUPAC Chemical Identifier Project, the project was launched at an IUPAC meeting in Washington, DC in March 2000. NIST led the technical development with key contributions from Stephen Stein, Stephen R. Heller, and Dmitrii Tchekhovskoi and the IUPAC CCINS committee oversaw it. In 2009, the InChI Trust was established as a UK-based charity to oversee future development.[83] Today, InChI is a widely adopted standard in cheminformatics and is integrated into major databases such as PubChem.[84] It can be generated through various software tools, including the official InChI software provided by the InChI Trust, as well as by standard libraries such as `OpenBabel`[85] and `RDKit`.[86]

The InChI workflow comprises three primary steps: normalization, canonicalization, and serialization. The input molecular structure is provided in CT file format. During normalization, the molecule is standardized using predefined rules. Standardization addresses variations in aromaticity (Kekulé vs. aromatic), tautomeric forms, protonation states, and charge distributions. These rules ensure conversion of diverse representations into a single, consistent form. Canonicalization, based on the Morgan algorithm,[31] assigns unique atom numbering to the standardized structure, similar to canonical SMILES. Serialization then converts the canonicalized molecular graph into a linear string representation.[39]

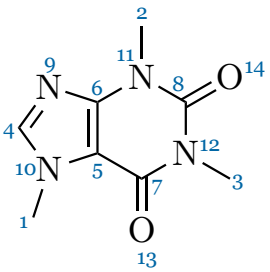
The final InChI string consists of multiple layers and sublayers, each encoding specific molecular information. The standard InChI includes four layers: the main layer, the charge layer, the stereochemistry layer, and the isotopic layer. It begins with the prefix "InChI=1S/". Non-standard InChI representations, which include additional layers, start with "InChI=1/". The main layer contains the chemical formula, a connectivity sublayer "/c" describing atom connections, and a hydrogen sublayer "/h" indicating hydrogens attached to heavy atoms. Figure 2.5 illustrates the InChI representation of caffeine. After normalization, canonicalization assigns unique atom labels (highlighted in blue). The connectivity sublayer "/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2" traces the molecular path using canonical labels. Brackets denote branches or ring closures, such as the ring formed by atoms 4, 5, 6, 9, and 10 in "10-4-9-6-5(10)". The hydrogen sublayer "/h4H,1-3H3" specifies that atom 4 has one hydrogen, while

Table 2.1: Examples of SMILES and InChI representations for various molecules.

Name	Structure	SMILES	InChI
Methane		C	InChI=1S/CH4/h1H4
Ethene		C=C	InChI=1S/C2H4/c1-2/h1-2H2
Hydrogen cyanide		C#N	InChI=1S/CHN/c1-2/h1H
Cyclohexane		C1CCCCC1	InChI=1S/C6H12/c1-2-4-6-5-3-1/h1-6H2
Benzene		c1ccccc1	InChI=1S/C6H6/c1-2-4-6-5-3-1/h1-6H
L-Alanine		C[C@@H](C(=O)O)N	InChI=1S/C3H7NO2/c1-2(4)3(5)6/h2H,4H2,1H3,(H,5,6)/t2-/m1/s1
D-Alanine		C[C@H](C(=O)O)N	InChI=1S/C3H7NO2/c1-2(4)3(5)6/h2H,4H2,1H3,(H,5,6)/t2-/m0/s1
Cis-Dichloroethylene		C(=C\Cl)\Cl	InChI=1S/C2H2Cl2/c3-1-2-4/h1-2H/b2-1-
Trans-Dichloroethylene		C(=C/Cl)\Cl	InChI=1S/C2H2Cl2/c3-1-2-4/h1-2H/b2-1+

atoms 1–3 each have three hydrogens. Bond orders are inferred from valences and hydrogen counts rather than explicitly stated.[39]

Double bond stereochemistry is specified in the stereochemistry layer using the sublayer "/b" with canonical atom labels, as illustrated in Table 2.1 for cis- and trans-dichloroethylene. The bond between atom 2 and 1 is denoted as cis ("/b2-1-"), while the trans configuration is represented by "/b2-1+". Tetrahedral stereochemistry is encoded in the stereochemistry layer with sublayers "/t", "/m", and "/s", as demonstrated by L- and D-alanine in Table 2.1. Here, "/t2-" designates atom 2 as a stereocenter with parity "-". The parity has no relation to R/S nomenclature but represents the canonical atom ordering. The sublayer "/m1" or "/m0" translates parity into absolute configuration (1 for L-alanine, 0 for D-alanine). "/s" clarifies how stereochemical information is provided: "/s1" indicates absolute stereochemistry, "/s2" signifies relative stereochemistry, and "/s3" denotes a racemic mixture of stereoisomers.[39]



InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3

Figure 2.5: Caffeine molecule and canonical order of atoms used in the InChI representation, along with the corresponding InChI string. The canonical order of atoms is determined during the canonicalization step of the InChI generation process. Serialization follows to convert the canonicalized molecular structure into the linear InChI string representation. The connectivity sublayer starting with "/c" shows this path through the molecule as it follows the canonical labels of each atom (highlighted in blue). The hydrogen layer indicates that atom 4 has one hydrogen atom and atoms 1–3 have three hydrogen atoms attached.

Until recently, the InChI format was significantly limited when it came to inorganic and organometallic compounds. Originally designed with organic chemistry in mind, the InChI algorithm traditionally severed metal–ligand bonds by treating them as ionic interactions. While this approach was effective for organic molecules and classical ions, it presented challenges for inorganic, coordination and organometallic compounds. In these cases, metal–ligand connectivity often carries essential stereochemical information that was lost when bonds were disconnected. This identifier also struggled to accurately represent hapticity in complexes such as ferrocene. However, in 2025, version 1.07 of the InChI software was released, introducing major improvements to address these issues. New rules were implemented to normalize and preserve metal–ligand bonds in coordination and organometallic compounds. These 'disconnection rules', based on coordination numbers and electronegativity differences, help determine which bonds to retain and which to remove as they are likely to be ionic. Support for the stereochemistry of inorganic compounds will continue to be developed in the future.[87]

Other notable molecular string representations exist, many of which are not widely used anymore. Examples include WLN[34,88,89] (1949), ROSDAL[38] (1989), and SLN[37,90] (1997). More recently, self-referencing embedded strings (SELFIES) has been introduced as a robust molecular string representation designed for ML applications.[91] SELFIES builds upon the SMILES notation but incorporates an underlying grammar that guarantees the generation of valid molecular structures, which is particularly advantageous for generative models in ML. Group SELFIES extends this concept by allowing the representation of larger chemical groups as single tokens, which in theory allows for representation of non-central chirality, something that all the discussed string notations cannot represent.[92] Although not a line notation, the MolAssembler graph-based representation is still worth mentioning, as it uses graph isomorphism to uniquely represent molecular structures including transition metal complexes and their stereochemistry.[93]

2.3 Graph Theory in Cheminformatics

As discussed previously, line notations such as SMILES[35,36] and InChI[39] are generated from the Ctab representation of molecules. The Ctab representation models molecules as mathematical graphs, with atoms as vertices and bonds as edges. Graph theory provides algorithms that are essential for cheminformatics, including those used to serialize these graphs, which leads to one-dimensional line notations. Furthermore, these algorithms facilitate ring perception, enabling the identification of cyclic substructures in molecules, which is a prerequisite for generating SMILES and InChI. This section briefly introduces graph theory. A graph $G(V, E)$ is defined by a set of vertices $V = \{i, j, k, ...\}$ and a set of edges $E = \{(i, j), (j, k), ...\}$. Vertices are connected by edges. If an edge (i, j) connects vertices i and j, both vertices are termed "adjacent". The degree of a vertex $\deg(i)$ refers to the number of edges connected to it. In chemical terms, vertices represent atoms, and edges represent bonds. Thus, the number of elements in $V(G)$ equals the number of atoms, while the number of elements in $E(G)$ equals the number of bonds in the molecule. The degree of a vertex $\deg(i)$ corresponds to the chemical concept of a coordination number. Edges may be directed or undirected. In an undirected graph, information flows both ways between two vertices. In contrast, directed edges have a specific orientation from one vertex to another, such that a path can only be traversed in one direction. Edges can also be assigned weights, which may represent bond orders or other properties. In different contexts, weights can denote distances, costs, or capacities. Graphs can be conveniently represented as matrices. Using a matrix representation allows the application of standard linear algebra operations to address graph-theoretical problems. One such matrix representation is the adjacency matrix. An adjacency matrix $\mathbf{A}_{\text{adj}}$ for an n_{atoms} molecule is an $n_{\text{atoms}} \times n_{\text{atoms}}$ matrix where the elements indicate the connectivity of individual vertices (atoms). If an element a_{ij} of the matrix equals one, vertices i and j are adjacent; otherwise, if a_{ij} equals zero, they are not adjacent. For an unweighted graph, the adjacency matrix is defined as follows:[94]

$$\mathbf{A}_{\text{adj}} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ a_{n1} & \dots & \dots & a_{nn} \end{bmatrix} \text{ with } \quad a_{ij} = \begin{cases} 1 & \forall (i,j) \in E(G) \\ 0 & \forall (i,j) \notin E(G) \end{cases} \tag{2.13}$$

In the following, three fundamental graph algorithms relevant to cheminformatics applications are presented: Dijkstra's algorithm[95] for shortest path calculations starting from a single source vertex, the Floyd-Warshall algorithm[96,97] for shortest path calculations between all pairs of vertices, and the breadth-first search (BFS) algorithm for ring perception. All algorithms are discussed for unweighted, undirected graphs, as is common in cheminformatics applications.

In graph theory, the distance between two vertices is determined by the smallest number

of edges connecting them. The shortest path between vertices i and j represents the least number of edges required to travel from i to j. Dijkstra's algorithm is widely used to find such shortest paths in graphs with non-negative edge weights.[95,98] Its utility spans multiple domains, from chemical applications to infrastructure planning, such as road network design and improving ship fuel efficiency. The algorithm proceeds by iteratively examining vertices connected by edges. An update to the original algorithm is characterized by a time complexity of $O(E+V \log V)$.[99,100] The original algorithm's logic is outlined in pseudocode in Algorithm 1, which scales with $O(V^2)$.[95,100]

Algorithm 1 Dijkstra's Algorithm

procedure DIJKSTRA($G(V, E)$, $start$)
    Initialize $distance[v] \leftarrow \infty$ for all $v \in V$
    Initialize $visited[v] \leftarrow false$ for all $v \in V$
    $distance[start] \leftarrow 0$
    $queue \leftarrow V$
    while $queue$ is not empty **do**
        $current \leftarrow$ vertex with minimum distance in $queue$
        Remove $current$ from $queue$
        $visited[current] \leftarrow true$
        for all neighbors $neighbor$ of $current$ **do**
            if not $visited[neighbor]$ **then**
                $newDistance \leftarrow distance[current] + 1$
                if $newDistance < distance[neighbor]$ **then**
                    $distance[neighbor] \leftarrow newDistance$
                end if
            end if
        end for
    end while
    return $distance$ array
end procedure

Initially, distances from the source vertex to all others are set to infinity, except the distance from the source to itself, which is zero. All vertices are initially marked as unvisited. A queue is created initially containing all vertices in the graph. In each iteration, the unvisited vertex with the smallest tentative distance is chosen from the queue. On the first iteration, this vertex is the starting vertex. The chosen vertex is then marked as visited and removed from the queue. For each of its adjacent unvisited vertices, the distance is recalculated by adding one edge to the distance from the source vertex to the current vertex. If the path through the current vertex yields a shorter distance, the recorded distance is updated. The process repeats until the queue is empty, ensuring the shortest paths from the source to all other reachable

vertices are found.[95]

For scenarios requiring shortest paths between every pair of vertices, the Floyd-Warshall algorithm is preferred. This algorithm is efficient for dense graphs and operates with a time complexity of $O(V^3)$.[101] It dynamically updates a distance matrix, initially populated with direct distances between all vertex pairs. The pseudocode for the Floyd-Warshall algorithm is provided in Algorithm 2.[96,97]

Algorithm 2 Floyd-Warshall Algorithm

```
procedure FLOYD-WARSHALL(G(V, E))
    Initialize distance[i][j] ← ∞ for all i, j ∈ V
    for all v ∈ V do
        distance[v][v] ← 0
    end for
    for all edges (u, v) ∈ E do
        distance[u][v] ← 1
        distance[v][u] ← 1                          ▷ For undirected graphs
    end for
    for all k ∈ V do
        for all i ∈ V do
            for all j ∈ V do
                if distance[i][j] > distance[i][k] + distance[k][j] then
                    distance[i][j] ← distance[i][k] + distance[k][j]
                end if
            end for
        end for
    end for
    return distance matrix
end procedure
```

The algorithm begins by creating a distance matrix where all pairwise vertex distances are initially set to infinity. For each existing edge, the distance between its connected vertices is updated to one, indicating a direct connection. For each vertex k, the algorithm evaluates whether the path from vertex i to vertex j via k is shorter than the currently recorded distance. If a shorter path is found, the distance matrix is updated. Once all vertices have been processed, the matrix contains the length of the shortest path between every pair of vertices.[96,97] The shortest path can also be stored, which can be written as a set of edges $p^s(i,j) = \{(i,k),(k,l),...,(m,j)\}$ connecting vertices i and j. From the graph distance d_{ij} between atoms i and j, the entire distance matrix $\mathbf{D} \in \mathbb{N}^{n_{\text{atoms}} \times n_{\text{atoms}}}$ can be constructed, where each element d_{ij} represents the graph distance between atoms i and j.

The BFS algorithm is a fundamental graph traversal method used to explore all vertices and edges of a graph systematically. It is particularly useful for identifying connected components and detecting cycles within graphs, which is essential for ring perception in chemical structures. The algorithm operates with a time complexity of $O(V + E)$.[102] Many ring perception algorithms in cheminformatics are based on BFS.[103–105] The pseudocode for the BFS algorithm is outlined in Algorithm 3.

Algorithm 3 Breadth-First Search (BFS) Algorithm for Cycle Detection

```
1: procedure BFS(G(V, E), start)
2:     Initialize visited[v] ← false for all v ∈ V
3:     Initialize parent[v] ← null for all v ∈ V
4:     Initialize queue ← {start}
5:     visited[start] ← true
6:     while queue is not empty do
7:         current ← Dequeue from queue
8:         for all neighbors neighbor of current do
9:             if not visited[neighbor] then
10:                visited[neighbor] ← true
11:                parent[neighbor] ← current
12:                Enqueue neighbor to queue
13:            else if neighbor ≠ parent[current] then
14:                Cycle Detected
15:            end if
16:        end for
17:    end while
18: end procedure
```

The algorithm starts by initializing a visited array to track which vertices have been explored. A queue is created, starting with the initial vertex. The starting vertex is marked as visited. The algorithm enters a loop that continues until the queue is empty. In each iteration, the vertex at the front of the queue is dequeued and processed. For each of its adjacent vertices, if they have not been visited, they are marked as visited and added to the queue. This process ensures that all vertices reachable from the starting vertex are explored in breadth-first order. A cycle is detected when a vertex that has already been visited is encountered during traversal and is not the parent of the current vertex. A chemically meaningful cycle is found when it corresponds to the smallest cycle for an atom in a molecular graph. This leads to a set of rings $C^s = \{C_1, C_2, ..., C_m\}$, where each ring is a set of edges $C = \{(i, j), (j, k), ..., (n, i)\}$ forming a closed loop in the molecular graph.

3 Motivation

Previous sections highlight that molecular identifiers such as InChI and SMILES were designed primarily for cheminformatics. These representations rely on Molfiles, where connectivity, bond order, and stereochemistry are explicitly predefined.[33] In quantum chemistry, however, structures often originate from methods that generate new stationary points without manual interference through conformational sampling,[48,49] reaction discovery[50,53,106,107] or transition state searches.[108–110] Within the Born–Oppenheimer approximation, each structure is defined exclusively by 3D atomic coordinates, with properties like connectivity inferred from solving the electronic Schrödinger equation rather than predefined rules.[66]

Classifying the vast number of generated structures, whether as conformers or distinct isomers, demands a molecular line representation. Raw 3D coordinates in XYZ files are inadequate due to their lack of translational, rotational, and permutational invariance, which allows coordinates of conformers to vary while representing identical structures. Normalized or sorted internal coordinates do not solve this problem, as they continue to change during conformational rearrangements. While converting coordinates to Molfiles enables SMILES or InChI generation, this approach introduces inconsistencies. Bond existence and order, derived from electronic structure, depend on the method used, since no quantum operator defines bonds objectively.[111] Different tools produce divergent results, complicating standardization. Moreover, most cheminformatics tools are based on valences to convert coordinates to integer bond orders, which are then used to generate InChI or SMILES. This approach is particularly problematic when applied to regimes of quantum chemistry, where valences can be violated, such as in transition states or nanoreactor-type simulations.[50]

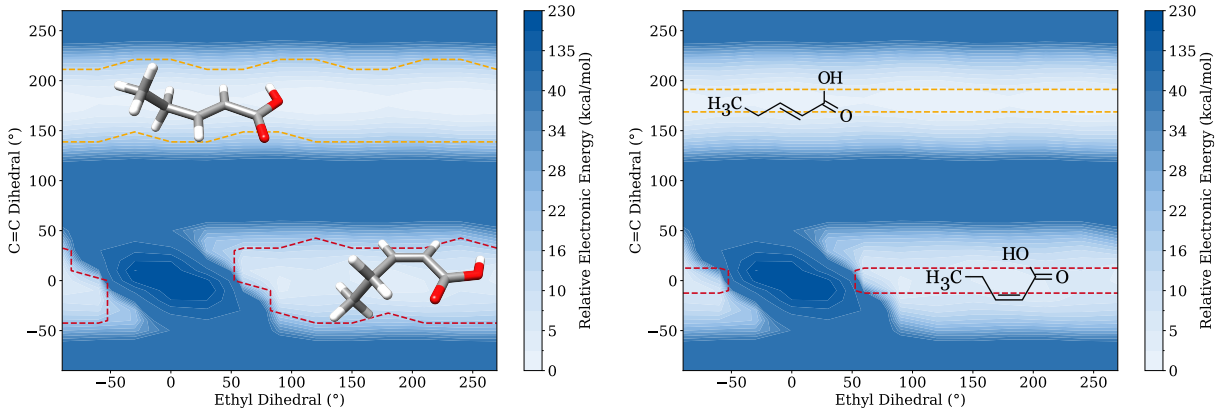
A novel identifier concept is required to address these challenges. This identifier is designed to meet the specific requirements of quantum chemistry and is optimized for use in quantum chemical databases. It must derive molecular identity directly from 3D coordinates without relying on valence rules or external tools and their subjective interpretations of bonds. While such an approach inevitably incorporates its own interpretation, it ensures consistency for a given coordinate set. The primary objective is to classify stationary points (for now, minima only) on the PES. Each distinct minimum should correspond to a unique molecular identity. Meanwhile, conformer structures that interconvert through low barriers should be represented by a single identifier. This *ansatz* distinguishes between conformers and isomers based on the energy barriers present on the PES, aligning with the refined IUPAC definition of isomers proposed by Poppe and Nogradi.[112] According to their interpretation, isomers are molecules with identical formulas that can be isolated, directly linking isomerism to the energy landscape

of a molecule. To establish a practical distinction between conformers and isomers, an energy barrier threshold of 20 kcal/mol is adopted as a central heuristic.[113] For viable deployment in quantum chemical workflows, the computational cost to generate the identifier must be comparable to, or lower than, a single-point semiempirical electronic structure calculation. Such strict efficiency constraints preclude comprehensive PES exploration during identifier generation. Instead, the 20 kcal/mol threshold functions as a conceptual guideline throughout the subsequent analyses, applied as a loose criterion rather than a strictly computed boundary. This threshold informs the development of the new identifier and provides a benchmark for evaluating how well existing identifiers align with the desired PES mapping.

The following examples illustrate the requirements for a quantum chemistry-based molecular identifier. Each example employs an approximate PES calculated on a grid of two internal coordinates, using MLIP UMA-S[65] nuclear potential energy (in the Born–Oppenheimer approximation) computations. Since coordinates remain fixed and no relaxation occurs, the resulting reaction energies and activation barriers serve illustrative rather than exact purposes. For comparison, InChIs are generated for all grid structures using `OpenBabel` 3.1.0.[85] It should be noted that InChI generation relies heavily on the parser tool to derive connectivity and bond order from 3D coordinates. Different tools may produce different InChIs for the same coordinates due to varying interpretations when bond breaking or formation occurs. The desired PES mapping relies on 20 kcal/mol basins, which are determined by calculating the energy required to reach an arbitrary point on the PES starting from each minimum along the minimum energy reaction pathway (MERP). Any basins that overlap are then merged.

First, a straightforward example is presented in which the InChI aligns with the desired PES mapping. E/Z isomerism is a common type of stereoisomerism found in double bonds. Figure 3.1 illustrates this form of isomerism using 2-pentenoic acid. The internal coordinates chosen are the two C-C-C-C dihedrals: one spanning the double bond and the other involving the ethyl group. The rotational barrier separating E and Z isomers significantly exceeds the 20 kcal/mol threshold. This results in a clear distinction between the E and Z isomers on the PES in Figure 3.1a indicated by the red and orange basins. In contrast, the rotational barriers for conformers of the ethyl group remain below 20 kcal/mol. Consequently, all ethyl conformers are collectively grouped into the same basins. In this context, InChI nearly aligns with the desired PES mapping, as it effectively differentiates between E and Z isomers. The Lewis structure interpretations of InChI are shown in Figure 3.1b.

Tautomerism presents a contrasting scenario where InChI diverges from the desired PES mapping in quantum chemistry. Tautomerization involves bond cleavage and formation, resulting in altered connectivity. In general, bond dissociation can require several tens of kcal/mol to overcome.[114] In tautomerization, the energy barrier typically ranges from 10 up to 45 kcal/mol, depending on the mechanism, solvation, temperature and other conditions.[115–118] Figure 3.2 depicts the illustrative PES for the tautomerization of isobutyramide to its imidic acid form. The PES is constructed using the N-H bond distance and the amide C-N-H bond angle. The



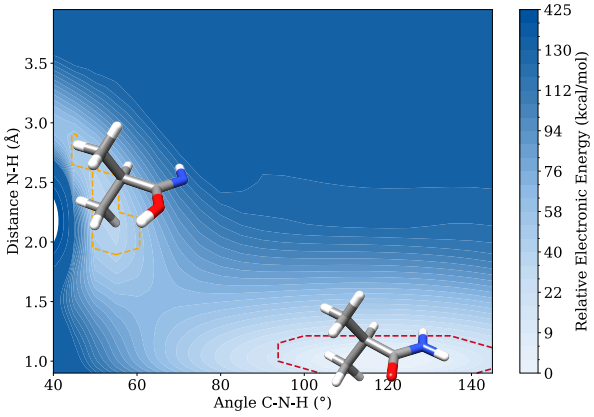
(a) Target PES mapping

(b) InChI-based PES mapping and InChI-based Lewis structures

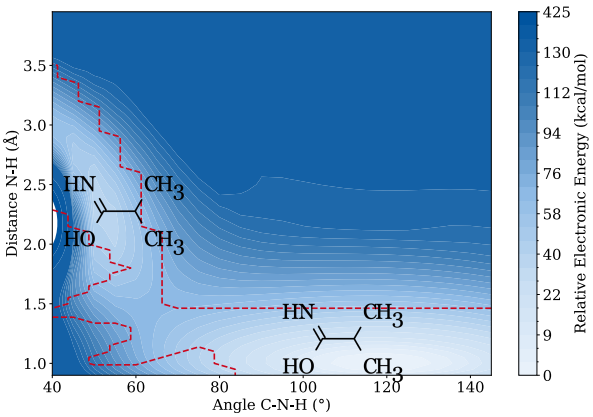
Figure 3.1: Illustrative PES for the E/Z isomerization of 2-pentenoic acid. Energy calculations utilize the MLIP UMA-S.[65] InChI identifiers are derived using `OpenBabel` 3.1.0.[85] The *E* and *Z* isomers manifest as distinct minima on the PES, separated by a substantial energy barrier. The target PES (a) mapping on the left side distinguishes between these isomers, whereas conformers associated with the ethyl group orientation are aggregated. The InChI PES mapping (b) similarly assigns unique identifiers to each isomer.

target PES mapping clearly differentiates between the two tautomers, separated by a barrier above 20 kcal/mol (Figure 3.2a). However, InChI defaults to treating tautomers as equivalent through tautomer normalization (see Lewis structures in Figure 3.2b), assigning identical identifiers to both forms. This design reflects InChI's intent to group tautomers that interconvert in solution within experimental time frames. While this aligns with the perspective of experimental chemists, quantum chemistry recognizes tautomers as distinct minima on the PES, each with unique properties. It has to be noted that a non-standard InChI layer, the fixed-H layer, can be used to distinguish tautomers; however, this layer is rarely applied in practice. SMILES in general does not apply tautomer normalization and thus does differentiate between tautomers.

Atropisomerism presents a unique challenge for molecular identifiers. Atropisomers emerge due to restricted rotation around a single bond, creating distinct stereoisomers.[113] Figure 3.3 shows the illustrative PES for a dibromo-biphenyl-thioamide derivative as an example of atropisomerism. The PES is constructed by varying the dihedral angles around the biphenyl and the thioamide C-N bond. Both rotations encounter substantial energy barriers: Biphenyl rotation is hindered by steric interactions between substituents, making it an atropisomeric axis. For dibromo-biphenyls, the barrier is estimated to be around 20 kcal/mol.[119] The thioamide C-N bond rotation is impeded by partial double-bond character resulting from resonance. Thioamides exhibit elevated bond orders, contributing to their high rotational barriers.[120] On the PES, all four atropisomers appear as separate minima, each separated by significant energy barriers. The desired PES mapping clearly distinguishes between these atropisomers (see Figure 3.3a). However, InChI cannot differentiate them, assigning identical identifiers to each form without any stereoinformation (see Figure 3.3b). In its current implementation,

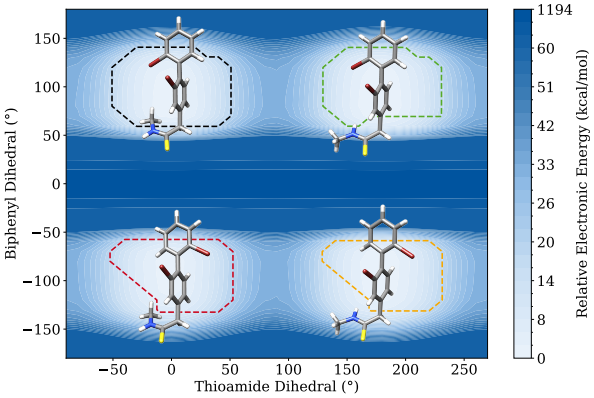


(a) Target PES representation

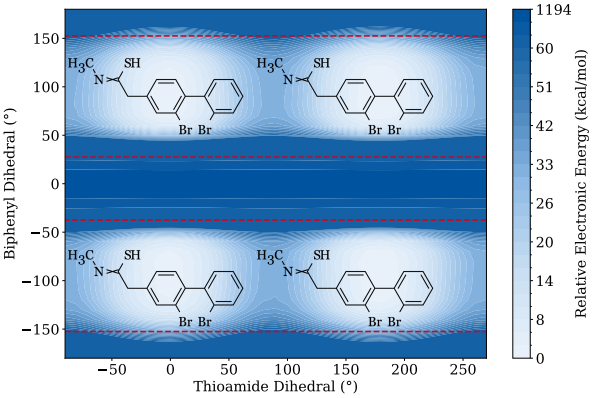


(b) InChI-based PES representation and InChI-based Lewis structures

Figure 3.2: Illustrative PES for the tautomerization of isobutyramide to its imidic acid form. Energies are computed with the MLIP UMA-S,[65] and InChI identifiers are created using `OpenBabel` 3.1.0.[85] The target PES mapping (a) differentiates between the two tautomers. However, the InChI PES mapping does not distinguish them, assigning identical InChI identifiers. This results from the standard InChI treating tautomers as equivalent by default through tautomer normalization (b). To obtain distinct InChI identifiers, the fixed-H layer would be necessary, though it is seldom applied.



(a) Desired PES mapping



(b) InChI-based PES mapping and InChI-based Lewis structures

Figure 3.3: Illustrative PES for the atropisomerism of a dibromo-biphenyl-thioamide derivative. Energy calculations utilize the MLIP UMA-S.[65] InChI identifiers are generated with `OpenBabel` 3.1.0.[85] All four atropisomers are distinct minima on the PES, separated by high energy barriers. The biphenyl rotation barrier arises from steric hindrance between bromine atoms. The thioamide C-N bond rotation barrier results from partial double-bond character due to resonance. The desired PES mapping distinguishes between all four atropisomers (a). In contrast, the InChI PES mapping assigns identical identifiers to each form (b).

InChI does not encode stereochemical information for single bonds, which is essential for representing atropisomerism. Furthermore, no existing molecular line notation aims to detect atropisomerism automatically. Some SMILES extensions can encode atropisomerism, but they do not attempt to identify it from the 3D structure. Instead, they rely on this stereochemical information being explicitly provided in the input structure, an assumption that does not hold for raw quantum chemical coordinates.[87,92] Other forms of non-tetrahedral chirality, such as helical chirality in helicenes and planar chirality in metallocenes, also remain unencoded by both InChI and SMILES. Fundamentally, these notations assign stereochemical information to individual atoms or bonds. While adequate for tetrahedral stereocenters, this localized approach fails to capture the holistic, group-based nature of non-tetrahedral chirality. To the best of the author's knowledge, Group-SELFIES is the only representation theoretically capable of encoding these alternative forms of chirality. However, for atropisomerism and other forms of non-tetrahedral chirality, it cannot automatically identify the relevant axis or assign the appropriate stereochemical tokens. Such automated detection is an essential requirement for any quantum chemistry-based identifier operating directly on raw 3D coordinates.[92] Beyond the challenges associated with atropisomerism, tautomer normalization in InChI converts the single bond in the thioamide to a double bond, which is inconsistent with the actual structure on the PES.

Cheminformatics has traditionally focused on organic chemistry over recent decades. While SMILES and InChI effectively describe two-center two-electron bonds in organic systems, they encounter limitations with η-bonds. Extensions to SMILES support basic molecular inorganic structures.[82,121,122] These extensions cover single-coordinating ligands and common metal geometries, including square planar, tetrahedral, and octahedral arrangements, along with their stereochemistry. Until recently, InChI treated metal centers and ligands as separate entities.[39] A recent development version of the InChI addresses molecular inorganics and allows normalization of metal-ligand bonds without cutting them.[87] However, stereochemistry for molecular inorganics remains unencoded in this version of InChI.

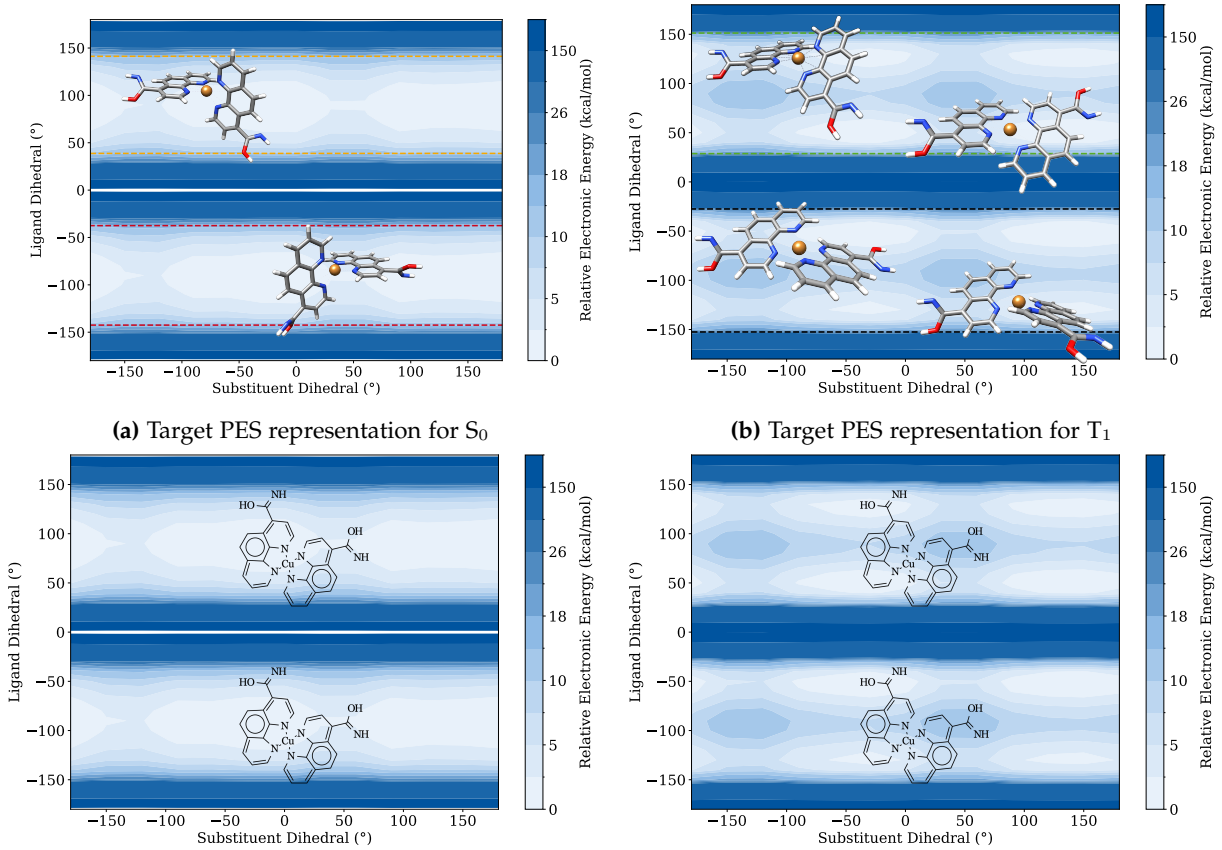
Further, most cheminformatics approaches are confined to ground-state chemistry. Yet, there are different PES for different electronic states of a molecule (e.g., singlet ground state vs. triplet excited state). A minimum on the ground state PES may not correspond to a minimum on the excited state PES, and vice versa. Activation barriers might also differ between states, and what constitutes conformers in one state may differ in another. Figure 3.4 shows the illustrative PES for the singlet ground and first triplet excited states of a copper(I) complex with two phenanthroline ligands. Copper(I) complexes with tetrahedral geometry tend to adopt a square planar configuration in the triplet excited state due to formal oxidation to copper(II).[123,124] The PES is constructed using the dihedral angle between both ligands and the imidic acid substituent dihedral angle. In the singlet ground state (S_0), the minimum corresponds to a pseudotetrahedral geometry. Both ligand orientations, which are enantiomers, are separated by a high energy barrier when coplanar due to steric hindrance. The target PES mapping distinguishes between the two tetrahedral enantiomers in the ground state (see

Figure 3.4a). In the first triplet excited state (T_1), the minimum shifts to a near square planar geometry. Ligand arrangements in this state can be either *cis* or *trans*, separated by a low barrier below 20 kcal/mol, indicating possible conformational interconversion. Due to the non-planarity of both ligands in *cis* and *trans* arrangements, there are two *cis* and two *trans* enantiomers. As the barrier between the *cis* and *trans* arrangements is low, they are considered conformers and grouped into the same basin on the PES. This yields two basins on the T_1 PES: one for one enantiomeric side (positive ligand dihedral angle) with both *cis* and *trans* arrangements, and another for the other enantiomer side (negative ligand dihedral angle) with both *cis* and *trans* arrangements as well. With different geometries in the ground and excited states, a different identifier is desired for each state, reflecting the distinct molecular shape and identity. The InChI PES mapping, however, does not differentiate these geometries, as stereochemical details for coordination complexes are not yet encoded in the current InChI preview for molecular inorganics.

The examples above expose fundamental limitations in existing molecular identifiers when applied to quantum chemistry. To address these challenges, a novel identifier is proposed: the Molecular Barcode (MolBar). This identifier must satisfy several core requirements derived from the PES-based perspective of computational chemistry: The first goal is to classify minima on the PES, focusing on distinguishing isomers while grouping conformers. A loose threshold of 20 kcal/mol serves as a guiding principle for differentiating between conformers and isomers, though it is not a strict rule due to the complexity of the PES and the associated computational cost. Second, the identifier must describe the 3D structure of molecules because the PES is inherently defined in 3D space, and geometries can vary across different electronic states. Third, when describing the 3D structure, the identifier must be invariant under translations, rotations, and permutations of atoms to ensure that the same structure yields the same identifier regardless of its orientation or atom ordering. Fourth, molecular identity must be extracted directly from 3D coordinates without using external tools or relying on valences to determine connectivity and bond order. Not relying on external tools ensures consistency because different tools may interpret the same coordinates differently, particularly in cases of bond breaking or formation. Omitting valence rules allows the identifier to be applied to a wide range of molecular systems, including those with non-standard bonding situations. These systems include transition states, organometallic complexes, coordination compounds, and potentially even unknown chemical species. This makes the identifier useful for exploring chemical space. Fifth, the identifier must differentiate constitutional isomers, stereoisomers, and, by default, prototropic tautomers, explicitly accounting for hydrogen positions. All stereochemical forms must be encoded, including non-tetrahedral chirality, such as axial and planar chirality, including atropisomerism. Finally, the computational cost of generating the identifier must be lower than or comparable to that of a single-point electronic structure calculation, ensuring its practical applicability in quantum chemistry workflows.

The subsequent sections develop Molecular Barcode (MolBar) systematically. First, foundational concepts are established, leading to formal mathematical definitions (Chapter 4). These

definitions are then translated into practical algorithms for computing MolBar directly from 3D coordinates (Chapter 5). The final section validates the identifier against the requirements outlined here (Chapter 6).



(a) Target PES representation for S_0

(b) Target PES representation for T_1

(c) InChI-based PES representation for S_0 and InChI-based Lewis structures

(d) InChI-based PES representation for T_1 and InChI-based Lewis structures

Figure 3.4: Illustrative PES for the singlet ground and first triplet excited states of a copper(I) complex with two phenanthroline ligands. Energy calculations were performed using the MLIP UMA-S.[65] InChI identifiers were generated via the InChI web preview for molecular inorganics, utilizing Molfiles from `OpenBabel` 3.1.0[85] as input. (a) The ground state (S_0) PES shows a minimum at a tetrahedral geometry, which exists as a pair of enantiomers. (b) The first triplet excited state (T_1) PES features minima with near-square planar geometries. Here, *cis* and *trans* arrangements are separated by low barriers, facilitating conformational interconversion. Because the structures are not perfectly planar, the two *cis* but also the two *trans* configurations form enantiomeric pairs, which are separated from each other by a high energy barrier at the strictly coplanar transition state. (c+d) The InChI-based PES mapping fails to differentiate these geometries on either the S_0 or T_1 surfaces, as stereochemical information for coordination complexes is currently not supported by the InChI preview for molecular inorganics.

4 Theoretical Derivation of MolBar

Sections 4.1 to 4.3 of this chapter are based on the following publication:

- van Staalduinen, N., Bannwarth, C., MolBar: a molecular identifier for inorganic and organic molecules with full support of stereoisomerism, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.

The author of this thesis is also the primary contributor to the original publication, with the only other author being the supervising professor as corresponding author. While the core methodology remains unchanged, the author of this thesis has restructured, reformatted, rephrased and adapted these sections to ensure coherence with the overall narrative of this work. This transformation integrates the sections into a unified monograph rather than a compilation of individual articles. Significant conceptual additions and clarifications compared to the original publication have been made:

- Section 4.1: The discussion now places greater emphasis on the distinctions between the line representation *ansatz* derived from an adjacency matrix eigenvalue spectrum and the established molecular line notations, specifically SMILES[35] and InChI.[39]

- Section 4.2: The indicator function $\mathcal{I}_{ij}$ has been redefined to function as an operator applied to a Coulomb matrix.

  - The explanation has been expanded to provide clearer insight into its physical meaning.

  - Its mathematical formulation is now presented in the context of activation free energy barriers between minima on the PES.

  To enhance the logical progression of the content, the topography matrix and fragmentation sections from the original publication have been consolidated into a single section. Certain general mathematical expressions have been relocated to the general theory section on graphs (Section 2.3).

- Section 4.3: The original discussion has been rephrased to blend into the overall narrative of the thesis, with a more explicit focus on the limitations of the topography matrix in representing chirality. The need for a third matrix to encode absolute configuration has been stated directly.

Furthermore, the foundational concept for this identifier originated in the author's Master's thesis, completed in 2022 at RWTH Aachen University under the same supervisor. Titled "Development of a Biunique Representation of Achiral and Chiral Molecules Based on Spectral Representations for Use in High-Throughput-Screening Databases", the thesis first introduced the core idea of utilizing the eigenvalue spectrum of matrices as a molecular line representation. Since then, the concept has evolved substantially: only the topology matrix (Section 4.1) remains from the original work, whereas the topography and chirality matrices were entirely developed over the course of the present PhD research.

4.1 Topology: Molecular Connectivity

Describing the connectivity of atoms within a molecule is essential for differentiating between distinct molecular structures. As outlined in Section 2.2, standard cheminformatics line notations such as SMILES[35] and InChI[39] rely on graph-based representations of molecules, supplemented by stereochemical annotations. This topological approach enables the distinction between constitutional isomers, which are molecules possessing identical molecular formulas but differing in atomic connectivity.[125] Figure 4.1 illustrates a straightforward example of constitutional isomerism involving two silver complexes.

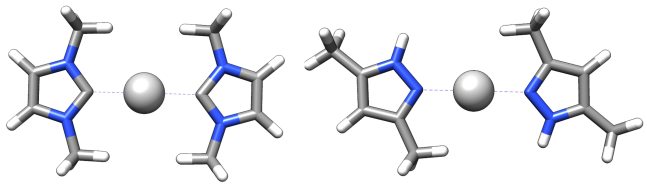


Figure 4.1: Two constitutional isomers of a silver complex, drawn from the tmQM dataset,[57] demonstrate varying connectivity patterns despite sharing the same molecular formula. Reprinted from van Staalduinen, N., Bannwarth, C., MolBar: a molecular identifier for inorganic and organic molecules with full support of stereoisomerism, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.

Section 2.3 introduces the mathematical framework of graphs, which can be used to represent molecular connectivity. Using the mathematical notation of graphs $G = (V, E)$, one method to encode molecular topology involves an adjacency matrix $\mathbf{A}_{\text{adj}} \in \mathbb{R}^{n_{\text{atoms}} \times n_{\text{atoms}}}$ as shown in Equation 2.13. However, adjacency matrices have limitations: their size grows quadratically with the number of atoms, and they lack permutation invariance regarding atomic ordering. Swapping atoms corresponds to rearranging the rows and columns of the adjacency matrix, leading to different representations for the same molecule. While not directly employing adjacency matrices, InChI and SMILES resolve this issue through canonicalization algorithms, ensuring a unique representation for a given molecular graph irrespective of the initial atomic order.[36,39]

However, theoretical chemistry offers a relevant analogy for representing molecular topology in an atomic-order-independent manner. Here, the Hückel matrix is a representation of π-conjugated systems, where diagonal elements reflect the relative energies of non-interacting p-orbitals, and off-diagonal elements describe interactions between neighboring p-orbitals.[13] The eigenvalues derived from the Hückel matrix correspond to molecular orbital energies under the LCAO approximation. Notably, the sorted eigenvalue spectrum remains invariant to permutations of atomic order.[126,127] This property eliminates the necessity for molecular graph canonicalization. Unlike SMILES and InChI, where each character in the line notation directly maps to a specific atom or bond, the eigenvalue spectrum does not establish a one-to-one relationship between individual eigenvalues and atoms. Instead, the eigenvalues encode topo-

logical substructures of the molecular graph, analogous to how canonical molecular orbitals are delocalized across the entire molecule rather than localized on specific atoms.

This insight leads to the adoption of the adjacency matrix spectrum for encoding molecular topology. This representation avoids the quadratic scaling of matrix elements with system size and can be expressed in linear notation. Yet, a basic adjacency matrix does not capture information about the specific types of atoms involved. Substituting one element for another leaves the matrix and its spectrum unchanged. To integrate atomic information, an extended adjacency matrix, as defined in Equation 4.1, is proposed. This formulation incorporates the atomic number Z_i and coordination number CN_i within the diagonal elements to characterize local atomic environments. The coordination number CN_i denotes the number of bonds associated with atom i and is equivalent to the graph degree $\deg(i)$ of the corresponding vertex in the molecular graph. When two atoms are bonded, the corresponding off-diagonal element is set to the arithmetic mean of their diagonal values. The elements of the topology matrix are specified as:

$$h_{ij}^{\text{topology}} := \begin{cases} (\mathrm{CN}_i + 1)Z_i & \forall i = j \\ \frac{1}{2}(h_{ii} + h_{jj}) & \forall i \neq j \wedge (i,j) \in E(G) \\ 0 & \forall i \neq j \wedge (i,j) \notin E(G) \end{cases}, \tag{4.1}$$

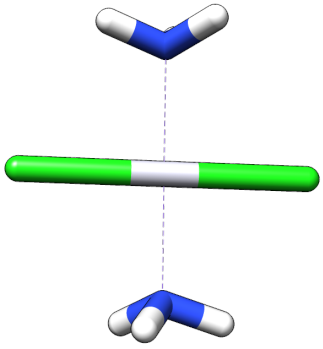
where h_{ii} and h_{jj} represent the diagonal elements for the i-th and j-th atoms, respectively. The eigenvalue spectrum $\sigma(\mathbf{H}^{\text{topology}}) \in \mathbb{R}^{n_{\text{atoms}}}$ is obtained by diagonalizing the molecular graph represented as $\mathbf{H}^{\text{topology}}$, which serves as an atomic-order-invariant description of the molecular topology.[126,127] To ensure the spectrum consists solely of integers, the eigenvalues are scaled by a factor of ten and rounded to the nearest whole number. Edge detection for the molecular graph and the implementation of the topology matrix are discussed in Section 5.2.

4.2 Topography: Relative Configuration

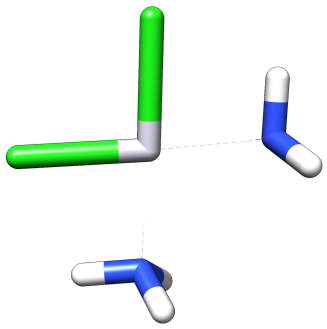
While molecular topology effectively distinguishes constitutional isomers, it does not fully encapsulate the intricacies of molecular structure. Stereoisomerism, for instance, occurs when molecules possess identical connectivity but vary in the spatial arrangement of their atoms. Stereoisomers are generally categorized into two types: diastereomers, which are not mirror images and display distinct physical properties, and enantiomers, which are non-superimposable mirror images and differ in specific properties such as optical rotation or circular dichroism, but are otherwise identical in many physical and chemical properties.

The concept of absolute configuration is explored in Section 4.3, while the current discussion centers on relative configuration, encompassing *cis*/*trans* isomers, E/Z isomers, and more intricate scenarios in coordination chemistry.[112] A well-known example of diastereomerism is

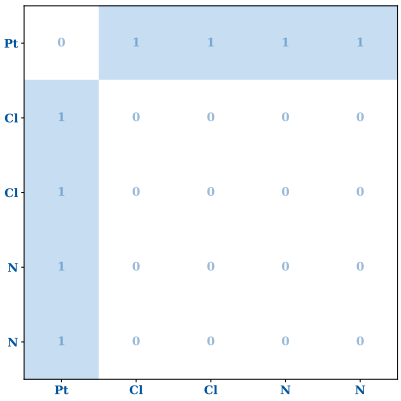
the *cis*/*trans* isomerism found in platinum complexes. Figure 4.2 depicts this phenomenon for two isomers of a platinum complex with two ammonia and two chloride ligands.



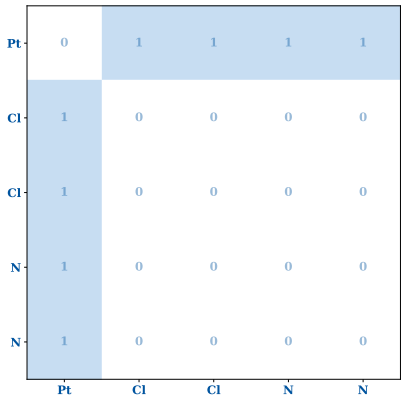
(a) Transplatin



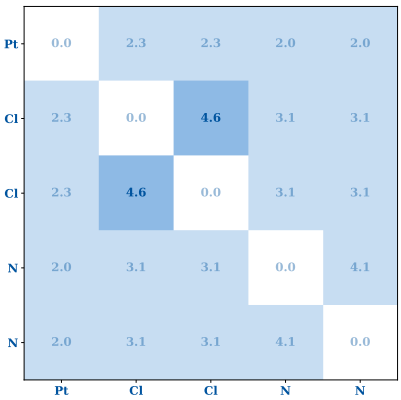
(b) Cisplatin



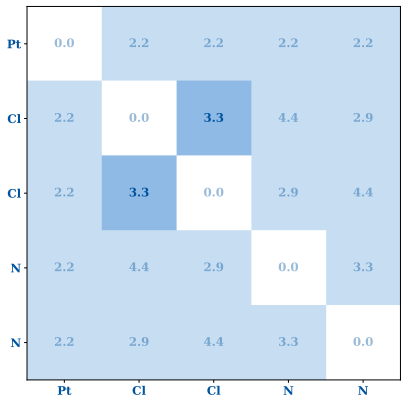
(c) Transplatin Adjacency matrix



(d) Cisplatin Adjacency matrix



(e) Transplatin Distance matrix



(f) Cisplatin Distance matrix

Figure 4.2: (a+b) *Cis*/*trans* isomers of a platinum complex with two ammonia and two chloride ligands. (c+d) The adjacency matrix is identical for both isomers due to their shared connectivity. (e+f) The distance matrix reveals different interatomic distances between the ligands (changing distance between chlorine atoms is highlighted), enabling differentiation between the diastereomers. For brevity, hydrogens are omitted in the matrices.

Both the *cis* and *trans* isomers (Figure 4.2a+b) exhibit the same adjacency matrix (Figure 4.2c+d), as their atomic connectivity remains unchanged. However, their spatial configurations differ, resulting in distinct interatomic distances. This distinction is reflected in the distance matrix

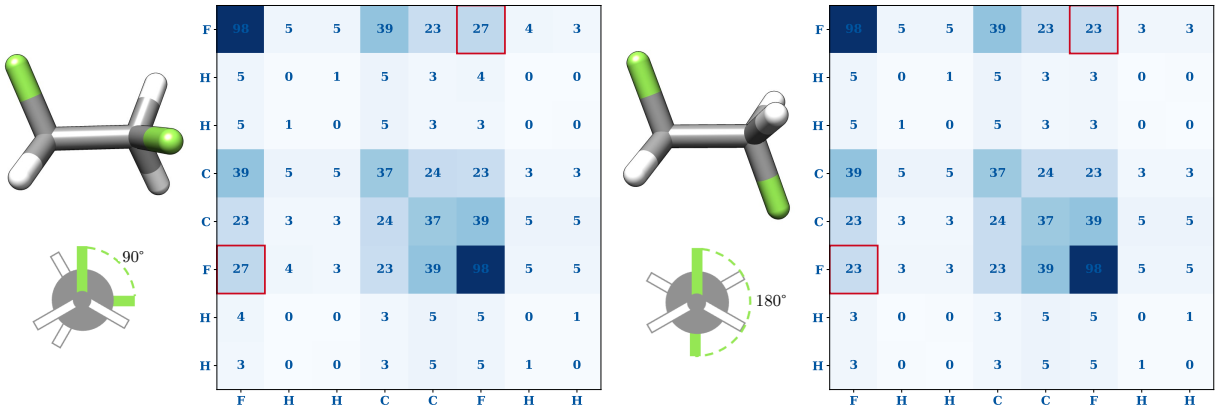
(Figure 4.2e+f), where the relative positions of the chlorine atoms lead to varying Cartesian distances between them. Unlike adjacency matrices, which use binary values (1 for bonded atoms, 0 otherwise), distance matrices contain real-valued entries representing the actual Cartesian distances between atom pairs. Since diastereomers differ in the relative positions of at least two atoms, their distance matrices will always differ for at least one pair of atoms, enabling their differentiation.

While distance matrices effectively describe relative configuration, they are not sufficient as standalone molecular identifiers. Distance matrices belong to a broader class of 3D representations, which have been widely used in ML to describe molecular shapes and correlate them with various properties.[128–130] Instead of using raw distances, many approaches employ the inverse of interatomic distances to model physical interactions, such as electrostatic forces, which are inversely proportional to distance. One such representation is the Coulomb matrix, introduced by Rupp et al. for predicting molecular atomization energies:[131]

$$b_{ij}^{\text{Coulomb}} := \begin{cases} 0.5 Z_i^{2.4} & \forall i = j \\ \frac{Z_i Z_j}{r_{ij}} & \forall i \neq j \end{cases}. \tag{4.2}$$

The diagonal elements represent a polynomial fit of atomic energies as a function of the atomic number Z_i. The off-diagonal elements incorporate the Coulomb repulsion operator with the inverse Cartesian distance r_{ij} between atom pairs, thereby enabling the characterization of molecular shape. Literature demonstrates the effectiveness of this matrix in describing constitutional isomers and diastereomers, utilizing its eigenvalue spectrum as a permutation-invariant descriptor of molecular geometry.[132] However, this representation lacks invariance across conformational space, rendering it unsuitable for use as a molecular identifier. Achieving a conformer-independent representation necessitates modifications to Equation 4.2. Conformational variations arise mainly from rotations about single bonds. These rotations lead to distinct Coulomb spectra, as the off-diagonal elements of the matrix in Equation 4.2 shift with changes in interatomic distance. Figure 4.3 illustrates this effect by showing how the off-diagonal matrix elements change for two conformers of 1,2-difluoroethane.

Nevertheless, the specific spatial orientation of fluorine atoms in 1,2-difluoroethane does not affect the molecule's identity, as transitions between conformers are rapid and reflect only conformational flexibility, not a fundamental change in molecular structure. To achieve conformer independence, a strategy is introduced that retains only those elements of the Coulomb matrix where the relative distances between atoms remain unchanged under rotations around single bonds or other low-energy transformations.



(a) Conformer 1 Coulomb matrix

(b) Conformer 2 Coulomb matrix

Figure 4.3: Conformational dependence of the Coulomb matrix in Equation 4.2[131] for two conformers of 1,2-difluoroethane optimized at the GFN2-xTB level of theory.[46] Rotation around the C-C single bond changes the distance between the two fluorine atoms, leading to different off-diagonal matrix elements (highlighted with a red box) and thus different eigenvalue spectra for the two conformers. For clarity, matrix elements are rounded and shown as integers. Modified and reprinted from van Staalduinen, N., Bannwarth, C., *MolBar: a molecular identifier for inorganic and organic molecules with full support of stereoisomerism*, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.

To formalize this observation in the context of the Coulomb matrix, an operator $\hat{\mathcal{I}}[\mathbf{B}]$ is defined that transforms an arbitrary Coulomb matrix $\mathbf{B}$ into a 3D representation excluding conformational information. This operator is applied as an element-wise product ($\circ$) between the Coulomb matrix $\mathbf{B} \in \mathbb{R}^{n_{\text{atoms}} \times n_{\text{atoms}}}$ and an indicator matrix $\boldsymbol{\mathcal{I}} \in \mathbb{R}^{n_{\text{atoms}} \times n_{\text{atoms}}}$, yielding $\mathbf{B}^{\text{topography}}$:

$$\mathbf{B}^{\text{topography}} := \hat{\mathcal{I}}[\mathbf{B}] = \boldsymbol{\mathcal{I}} \circ \mathbf{B} \tag{4.3}$$

The elements of $\mathbf{B}^{\text{topography}}$ are defined as:

$$b_{ij}^{\text{topography}} := \begin{cases} (\text{CN}_i + 1)Z_i & \forall i = j \\ \frac{1}{2}(b_{ii} + b_{jj})/R_{ij} & \forall i \neq j \wedge \mathcal{I}_{ij} = 1 \\ 0 & \forall i \neq j \wedge \mathcal{I}_{ij} = 0 \end{cases} \tag{4.4}$$

This matrix is termed the "topography" matrix, as it captures the 3D identity of a molecule while excluding conformational variations. Analogous to the topology matrix (Equation 4.1), the diagonal elements of $\mathbf{B}^{\text{topography}}$ incorporate the coordination number CN_i of each atom i and its atomic number Z_i to reflect element information per atom. The off-diagonal elements, derived from the Coulomb matrix, are retained only if the indicator matrix $\boldsymbol{\mathcal{I}}$ evaluates to 1 for the corresponding atom pair i and j; otherwise, they are set to zero. Thus, the operator $\hat{\mathcal{I}}[\mathbf{B}]$ transforms a given matrix $\mathbf{B}$ by applying the binary indicator function $\boldsymbol{\mathcal{I}}$, which determines whether each element of $\mathbf{B}$ should be retained or discarded. While a trivial choice of $\boldsymbol{\mathcal{I}}$ as a matrix of ones recovers the original (conformer-dependent) Coulomb matrix, setting $\boldsymbol{\mathcal{I}}$ to the identity matrix retains only diagonal elements, reducing the representation to atomic

properties and losing all 3D and topological information. Instead of directly using the Cartesian distances r_{ij}, the topography matrix uses modified R_{ij} distances. For an explanation, see the end of this section. The eigenvalue spectrum $\sigma(\mathbf{B}^{\text{topography}}) \in \mathbb{R}^{n_{\text{atoms}}}$ is obtained by diagonalizing $\mathbf{B}^{\text{topography}}$, which serves as an atomic-order-invariant description of the molecular topography.[126,127] To ensure the spectrum consists solely of integers, the eigenvalues are scaled by a factor of ten and rounded to the nearest whole number.

The construction of a meaningful indicator matrix $\boldsymbol{\mathcal{I}}$ draws on the kinetic perspective outlined in the Motivation (Chapter 3). Each distinct minimum, separated by a high energy barrier on the PES, is treated as a unique molecular identity. Conformer structures that interconvert via low barriers, however, share the same molecular identity. To implement this concept, a threshold activation free energy $\Delta G(i,j)^{\ddagger}_{\text{thrs.}}$ of 20 kcal/mol, following LaPlante et al.[113], is used for guidance. The following first develops a formal definition of $\boldsymbol{\mathcal{I}}$ in an ideal setting that assumes complete knowledge of the PES, before introducing the practical heuristic used to approximate it without exhaustive PES exploration.

In this idealized setting, where the complete PES and all its minima are known, the indicator function $\mathcal{I}_{ij}$ is defined for each pair of atoms i and j. Thus, the analysis must center on individual atom pairs rather than the molecule in its entirety. The central question revolves around whether the distance between atoms i and j constitutes a defining feature of molecular identity, or if it can vary due to low-energy conformational changes. To investigate this, the set of all known minima $\{\alpha, \beta, \ldots\}$ on the PES would be examined for a given elemental composition. Here, α denotes the input structure provided to MolBar, while β and others represent lower or higher energy minima. Each minimum corresponds to a distinct molecular geometry, characterized by an equilibrium interatomic distance $\{r_{ij}^{(\alpha)}, r_{ij}^{(\beta)}, \ldots\}$. The key consideration is: What is the minimum activation energy $\Delta G(i,j)^{\ddagger}$ required to alter the interatomic distance between atoms i and j by a significant margin ($|r_{ij}^{(\beta)} - r_{ij}^{(\alpha)}| > \delta$)? Significant displacement is essential, as some conformers involve only localized changes, leaving r_{ij} unaffected. In other words, the reaction or conformer coordinate must involve the relative movement of atoms i and j for the minimum to be included in the analysis. The activation free energy needed to achieve a significant change in the interatomic distance between atoms i and j would be determined by the energy barrier $G_{\alpha\beta}^{\text{TS}} - G^{(\alpha)}$ along an IRC connecting two minima α and β on the PES. From the set of activation energies, the minimum value $\Delta G(i,j)^{\ddagger}$ would be selected to represent the energy barrier for changing the distance between atoms i and j. This energy is formally expressed as:

$$\Delta G(i,j)^{\ddagger} = \min_{\substack{\beta \\ |r_{ij}^{(\beta)} - r_{ij}^{(\alpha)}| > \delta}} \left[G_{\alpha\beta}^{\text{TS}} - G^{(\alpha)} \right]. \tag{4.5}$$

A high barrier suggests that the distance is a defining characteristic, leading to $\mathcal{I}_{ij}$ being set to 1. Conversely, a low barrier indicates variability, resulting in $\mathcal{I}_{ij}$ being set to 0. Thus, $\Delta G(i,j)^{\ddagger}$ would serve as a measure to define the indicator function $\mathcal{I}_{ij}$:

$$\mathcal{I}_{ij} := \begin{cases} 1 & \text{if } i = j, \\ 1 & \text{if } i \neq j \wedge \Delta G(i,j)^{\ddagger} \geq \Delta G(i,j)^{\ddagger}_{\text{thrs.}}, \\ 0 & \text{if } i \neq j \wedge \Delta G(i,j)^{\ddagger} < \Delta G(i,j)^{\ddagger}_{\text{thrs.}} \end{cases} \tag{4.6}$$

However, directly computing $\Delta G(i,j)^{\ddagger}$ for all atom pairs i and j is computationally infeasible, as it would require exhaustive exploration of the PES for each molecular system. As a molecular identifier must be computationally efficient, an alternative heuristic approach is proposed to approximate $\mathcal{I}_{ij}$, motivated by the observations presented in Chapter 3. The PES of a molecule is usually described using internal coordinates q_k, such as bonds, angles, dihedrals, or combinations thereof. Chapter 3 highlighted several cases where internal coordinates changed with varying energy barriers, including: bond changes, as seen in tautomerization, angle changes, such as those between the S_0 and T_1 states of copper(I), and dihedral changes, including *cis*/*trans* isomerism and atropisomerism. For bond-breaking or bond-forming events, i.e., for two atoms i and j, where the graph distance is one ($d_{ij} = 1$), $\mathcal{I}_{ij}$ is assigned a value of 1. Such events typically involve significant energy barriers (such as bond dissociation or tautomerization) and alter the molecular identity. Similarly, if atoms are separated by two bonds ($d_{ij} = 2$) that form an angle between two bonds, $\mathcal{I}_{ij}$ is also set to 1. Equilibrium bond angles are generally well-defined in the ground state, and changes often require high-energy processes, such as electronic excitation as in the copper(I) example. Exceptions, like nitrogen inversion with its low barrier for angle change, are addressed in Section 5.5. Dihedral changes present a more nuanced scenario, as they can range from low-barrier conformational adjustments to high-barrier isomerization processes, such as atropisomerism. The rotation around a bond may be hindered by three primary factors: electronic effects (e.g., double bonds or conjugation), steric effects (e.g., bulky substituents), and ring effects (e.g., dihedrals in small rings). To account for these factors, a set of electronically and sterically hindered bonds M is defined (see Sections 5.3 and 5.4), along with the set of rings C^s (see Section 2.3). For atoms i and j separated by three or more bonds ($d_{ij} \geq 3$), $\mathcal{I}_{ij}$ is set to 1 if all bonds e_{kl} excluding the bonds directly connected to atoms i and j along the shortest path $p^s(i,j)$ belong to either the set of hindered bonds $e_{kl} \in M$ or a ring $e_{kl} \in C^s$. Otherwise, $\mathcal{I}_{ij}$ is set to 0. This means that for a double bond, the distance between the substituents is retained in $\mathbf{B}^{\text{topography}}$ to describe E/Z isomerism. Similarly, for atropisomers, where rotation around a single bond is hindered by bulky substituents, the relevant interatomic distances are preserved. For atoms separated by more than three bonds, the same logic applies. This is because if the distances between all intermediate atoms are preserved due to high energy barriers, the overall distance between atoms i and j is also preserved. In these regions, interatomic distances remain relatively static across the conformational landscape, as modifications would necessitate overcoming high energy barriers. This leads to the definition of rigid fragments within the molecule, where all interatomic distances are effectively fixed in conformational space. The heuristic definition of the indicator function $\mathcal{I}_{ij}$ is thus formalized as follows:

$$\mathcal{I}_{ij} = \begin{cases} 1 & \forall\, i = j \\ 1 & \forall\, i \neq j \wedge d_{ij} = 1 \ \#\text{Bonds} \\ 1 & \forall\, i \neq j \wedge d_{ij} = 2 \ \#\text{Angles} \\ 1 & \forall\, i \neq j \wedge d_{ij} \geq 3 \wedge \forall\, e_{kl} \in p^s(i,j) : e_{kl} \in (C^s \cup M) \ \#\text{Dihedrals} \\ 0 & \text{else} \end{cases} \tag{4.7}$$

Consequently, the indicator matrix $\boldsymbol{\mathcal{I}}$ is referred to as the rigidity matrix, as it delineates which atomic pairs belong to the same rigid fragment with fixed interatomic distances in conformational space.

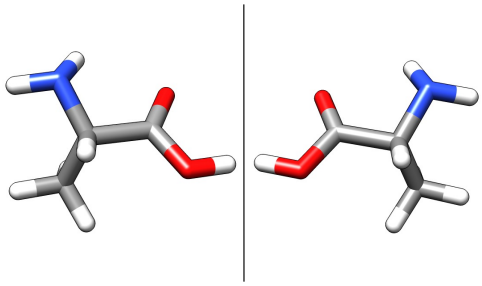
Nevertheless, even after applying the rigidity matrix $\boldsymbol{\mathcal{I}}$ to the Coulomb matrix, the resulting topography matrix $\mathbf{B}^{\text{topography}}$ is still affected by small variations, such as bond vibrations or minor changes in bond angles due to different computational methods used for optimizing the geometry. Therefore, a "structure idealization" or "structure unification" step follows after fragmentation. This idealization unifies the varying distances to a single representative value. This is achieved through a geometry optimization step, which idealizes the fragment structures to ideal values for bond lengths, angles, and dihedrals. Therefore, the topography matrix relies on R_{ij} as the idealized distance between atoms i and j after structure unification, rather than the actual Cartesian distance r_{ij}. This process is detailed in Section 5.5. After this step, the topography matrix $\mathbf{B}^{\text{topography}}$ becomes invariant to conformational changes and minor variations in geometry, thereby serving as a robust representation of the molecule's relative configuration.

4.3 Absolute Configuration: Chirality

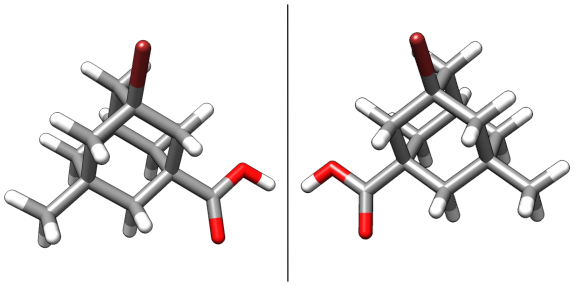
The topography matrix, as defined in Equation 4.4, encodes the relative configuration of a molecule through a set of overlapping rigid fragments with their respective relative configurations. However, it does not address the absolute configuration of the molecule. In a Coulomb-type matrix like the topography matrix, absolute configuration information is absent because interatomic distances remain unchanged under spatial reflection. Thus, to comprehensively describe a chiral molecule, both the topology and topography matrices must be supplemented by a third matrix.

Chirality holds significant importance in chemistry and biochemistry, as it profoundly affects molecular properties.[125] Enantiomers may display distinct pharmacological behaviors, differences in (bio)-chemical reactivity, and unique circular dichroism spectra.[125] Chirality commonly arises from central chirality at tetrahedral carbon atoms, as seen in alanine (Figure 4.4a). It can also emerge in systems where the chiral element is delocalized, such as in tetrasubstituted adamantane (Figure 4.4b).[125] Chirality can also originate from an axis such as

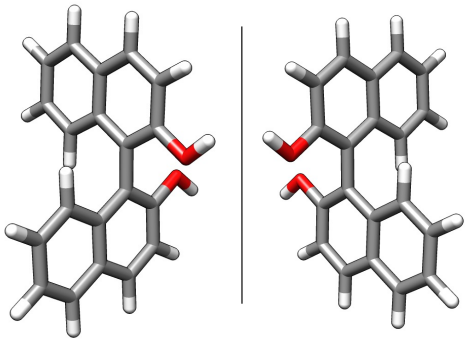
in atropisomerism, where atoms are arranged sequentially in either a clockwise or counter-clockwise manner around it (Figure 4.4c). Alternatively, it can arise from a plane, where atoms are positioned on opposite sides relative to a chiral plane (Figure 4.4d).[125] In all instances, chiral molecules display rigidity within the particular segment that defines their chirality— be it a chiral center, axis, or plane. This rigidity permits the application of the previously introduced concept of rigid fragments in Equation 4.7. Given that chirality only manifests in substructures that are rigid, it is adequate to describe the absolute configuration of these rigid fragments to characterize the absolute configuration of the entire molecule.



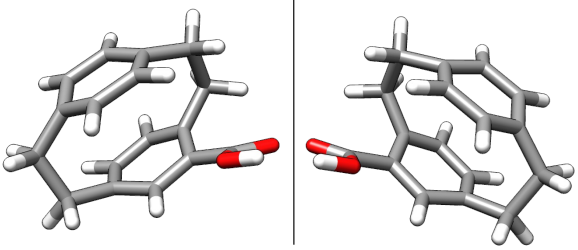
(a) alanine (central chirality)



(b) 3-bromo-5-methyladamantane-1-carboxylic acid (central chirality)



(c) [1,1′-Binaphthalene]-2,2′-diol (axial chirality, atropisomerism)



(d) 4-Carboxy[2.2]paracyclophane (planar chirality)

Figure 4.4: Examples of chiral molecules featuring different types of stereogenic units. Reprinted from van Staalduinen, N., Bannwarth, C., *MolBar: A molecular identifier for inorganic and organic molecules with full support of stereoisomerism*, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission of the Royal Society of Chemistry under the CC-BY 3.0 license.

Describing the absolute configurations of substructures, a fragment-based matrix is introduced, defined over the set of rigid fragments. This chirality matrix has dimensions $n_{\text{fragments}} \times n_{\text{fragments}}$, where $n_{\text{fragments}}$ represents the number of rigid fragments in the molecule. The chirality matrix is analogous to the topography matrix (Equation 4.4), but in fragment space rather than atomic space:

$$g_{ab}^{\text{chirality}} := \begin{cases} G_a p_a & \forall a = b \\ \frac{1}{2}(g_{aa} + g_{bb})/d_{ab} & \forall a \neq b \end{cases} \tag{4.8}$$

In this matrix, the diagonal elements represent the absolute configuration of individual fragments. The off-diagonal elements encode the relationships between chiral and achiral fragments. The variable $G_a \in \{-1, 0, 1\}$ serves as a chirality index, indicating whether a fragment a is achiral (0) or one of its two enantiomers (-1 or 1). An explicit definition of G_a is provided in Section 5.6.1. The parameter p_a denotes the priority of fragment a, analogous to the Cahn–Ingold–Prelog (CIP) rules but extended to entire fragments within the molecule. This parameter uniquely characterizes each fragment, and its detailed formulation is given in Section 5.6.2. Since in fragment space Cartesian distances are not defined, the graph distance d_{ab} is used instead, which counts the number of edges along the shortest path connecting fragments a and b in the fragment-based graph. Diagonalizing $\mathbf{G}^{\text{chirality}}$ produces the spectrum $\sigma(\mathbf{G}^{\text{chirality}}) \in \mathbb{R}^{n_{\text{fragments}}}$. To ensure the spectrum consists solely of integers, the eigenvalues are scaled by a factor of ten and rounded to the nearest whole number. This spectrum decomposes the absolute configuration of the molecule into contributions from its constituent fragments. This approach is not limited to atomic chiral centers but naturally extends to axial and planar chirality.

5 Practical Implementation

Sections 5.2 to 5.5 of this chapter are based on the following publications:

1. van Staalduinen, N., Bannwarth, C., MolBar: a molecular identifier for inorganic and organic molecules with full support of stereoisomerism, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.

2. van Staalduinen, N., Bannwarth, C., When Bonds Resist Rotation: Automated Classification and Representation of Atropisomers in the MolBar Identifier, *J. Chem. Inf. Model.*, **2026**, *66*, 10935–10955, DOI: `10.1021/acs.jcim.6c00688`, with permission from the American Chemical Society.

The author of this thesis is also the primary contributor to the original publications, with the only other author being the supervising professor as corresponding author. While the core methodology remains unchanged, the author of this thesis has restructured, reformatted, rephrased and adapted these sections to ensure coherence with the overall narrative of this work. This transformation integrates the sections into a unified monograph rather than a compilation of individual articles. Significant conceptual additions and clarifications compared to the original publication have been made:

- Section 5.2: In the original publication (1), the GFN-FF algorithm was implemented in MolBar to assign connectivity. Discussion of the algorithm used was limited, and readers were referred to the original GFN-FF publication. In this work, a method specifically tailored to MolBar was developed and compared with the GFN-FF topology algorithm and the Connect-The-Dots algorithm from `RDKit`. A comprehensive discussion of the results and limitations of each method was added.

- Section 5.3: In the original publication (2), the machine-learning model for predicting electronically hindered bonds was developed, trained and benchmarked. Here a detailed comparison with the `xyz2mol` implementation from `RDKit` has been added, including a discussion of advantages and disadvantages of both methods. Further, the computational cost of MolBar, `xyz2mol` and GFN2-xTB for bond assignment has been benchmarked and discussed.

- Section 5.4: Compared to the original publication (2), the discussion on atropisomerism has been reworded and integrated into the narrative of this work.

- Section 5.5: Compared to the original publication (1), the algorithm for assigning reference angles underwent a complete revision. The discussion was expanded to include cases of η-coordination when assigning reference angles. Originally, a standard optimizer from `SciPy` was used for structure unification. In this work, an optimizer was developed specifically for MolBar. This MolBar optimizer was compared with the `SciPy` Newton-CG optimizer and the approximate normal coordinate optimizer (ANCOpt) algorithm from CREST.

5.1 Outline of this Chapter

In Chapter 4, the theoretical foundation for a molecular identifier termed Molecular Barcode (MolBar) tailored to quantum chemistry applications was established. To render MolBar practically applicable, a comprehensive implementation strategy is presented in this chapter. The `molbar` implementation is distributed as an open-source `Python` package. Resource-intensive operations are handled by subroutines written in `Fortran`. Both the program package and its accompanying documentation are accessible at `https://git.rwth-aachen.de/bannwarthlab/molbar`. Installation is facilitated via the Python Package Index (PyPI) using the command `pip install molbar`.

At its core, MolBar employs a set of matrices. Specifically, it utilizes the rounded eigenvalue spectra of these matrices to encode the topology, topography, and absolute configuration of molecular fragments. The three matrices employed are as follows:

- the topology matrix $\mathbf{H}^{\text{topology}}$ (Equation 4.1),

$$h_{ij} := \begin{cases} (\text{CN}_i + 1)Z_i & \forall i = j \\ \frac{1}{2}(h_{ii} + h_{jj}) & \forall i \neq j \wedge (i,j) \in E(G) \\ 0 & \forall i \neq j \wedge (i,j) \notin E(G) \end{cases}, \tag{4.1}$$

- the topography matrix $\mathbf{B}^{\text{topography}}$ (Equation 4.4),

$$b_{ij} := \begin{cases} (\text{CN}_i + 1)Z_i & \forall i = j \\ \frac{1}{2}(b_{ii} + b_{jj})/R_{ij} & \forall i \neq j \wedge \mathcal{I}_{ij} = 1 \\ 0 & \forall i \neq j \wedge \mathcal{I}_{ij} = 0 \end{cases} \tag{4.4}$$

- the rigidity matrix $\boldsymbol{\mathcal{I}}$ (Equation 4.7),

$$\mathcal{I}_{ij} = \begin{cases} 1 & \forall, i = j \\ 1 & \forall, i \neq j \wedge d_{ij} = 1 \ \#\text{Bonds} \\ 1 & \forall, i \neq j \wedge d_{ij} = 2 \ \#\text{Angles} \\ 1 & \forall, i \neq j \wedge d_{ij} \geq 3 \wedge \forall\, e_{kl} \in p^s(i,j) : e_{kl} \in (C^s \cup M) \ \#\text{Dihedrals} \\ 0 & \text{else} \end{cases} \tag{4.7}$$

- as well as the chirality matrix $\mathbf{G}^{\text{chirality}}$ (Equation 4.8):

$$g_{ab} = \begin{cases} G_a p_a & \forall a = b \\ \frac{1}{2}(g_{aa} + g_{bb})/d_{ab} & \forall a \neq b \end{cases}. \tag{4.8}$$

Concluding those four matrices, the following properties are required to compute them:

- Z_i (the atomic number of atom i),
- CN_i (the coordination number of atom i),
- $E(G)$ (the set of edges in the molecular graph G),
- R_{ij} (the idealized distance between atoms i and j after structure unification),
- C^s (the set of rings in the molecular graph),
- $M_{\mathrm{electronic}} \cup M_{\mathrm{steric}}$ (the bonds that are rotationally hindered due to electronic and steric effects),
- G_a (the chirality index of rigid fragment a),
- p_a (the priority of rigid fragment a),
- $p^s(i,j)$ (the shortest path between atoms i and j in the molecular graph),
- d_{ij} (the graph distance between atoms i and j), and
- d_{ab} (the graph distance between rigid fragments a and b).

Certain quantities, such as the atomic number Z_i, are tabulated properties. Others, including the coordination number CN_i, graph distance d_{ab}/d_{ij}, shortest path $p^s(i,j)$, and the set of rings C^s, can be derived once the molecular graph $E(G)$ is established. These can be identified using standard graph-theoretical algorithms, as discussed in Section 2.3. While such algorithms are readily available in packages like `networkx`,[133] the graph algorithms in the `molbar` implementation are tailored specifically for MolBar and rely exclusively on `NumPy`[134] to eliminate external dependencies.

Other properties integral to MolBar necessitate the development of dedicated algorithms. These algorithms are optimized for performance and tailored to the specific requirements of MolBar. The development and benchmarking of these algorithms are discussed in the subsequent sections. The properties requiring specialized algorithms include:

- the set of bonds in the molecular graph $E(G)$ (Section 5.2),
- the set of electronically hindered bonds $M_{\mathrm{electronic}}$ (Section 5.3),
- the set of sterically hindered bonds M_{steric} (Section 5.4),
- the structure unification of rigid fragments to yield idealized distances R_{ij}(Section 5.5),
- the chirality index G_a for each rigid fragment, along with the priorities p_a (Section 5.6).

For each of these properties, dedicated sections are provided to detail their implementation within the `molbar` package. Each section outlines previous work from the literature, followed by a detailed description of the algorithms developed in this work specifically for MolBar. The algorithms are implemented in `Python` (only relying on `NumPy`[134] and `SciPy`[135]) and `Fortran`, with benchmarks comparing them to existing methods where applicable.

5.2 Topology Matrix: Determining Connectivity from 3D Coordinates

5.2.1 Introduction

Computing MolBar requires a reliable criterion for determining whether two atoms are bonded, as this information is needed to construct the topology matrix defined in Equation 4.1. In quantum chemistry, a persistent challenge arises from defining molecular connectivity from 3D atomic coordinates. This contrasts with conventional cheminformatics approaches, where connectivity is routinely extracted from 2D Lewis depictions or retrieved from CT file formats that store bond information explicitly. Because standard procedures for generating SMILES or InChI identifiers rely on the availability of a CT file and thus presuppose known topological data, these representations do not address the problem of deriving connectivity from 3D coordinates.[35,39] As elaborated in Section 2.2, quantum chemistry programs commonly encode molecular structures as lists of atomic coordinates and atomic numbers, without explicit bond information. Here, SMILES and InChI delegate the non-trivial inference of connectivity to an external tool.

One potential solution involves employing a connectivity parser to generate a CT file, which can subsequently be used to produce SMILES or InChI identifiers.[85,86,136] However, this approach is challenged by the lack of a universally accepted definition of a chemical bond within quantum mechanics. In the absence of such a criterion, determining whether two atoms are bonded is subject to interpretation, whether by the chosen method or through manual assessment.[137] Various methods exist to assign bonds, ranging from geometry-based approaches to analyses of electron density. Quantum chemists use the electronic structure of molecules to determine bonds. The most prominent methods partition the electron density into atomic contributions using Mulliken population analysis,[138] Wiberg,[139] or Mayer[140] bond order indices. Other techniques analyze the electron density itself, such as quantum theory of atoms in molecules (QTAIM) by Richard Bader[141] and electron localization function (ELF) by Becke and Edgecombe.[142] probability density analysis (PDA) by Lüchow and coworkers instead analyzes the modulus squared of the wave function $|\Psi|^2$, which remains 3N-dimensional rather than the 3-dimensional electron density.[143] While these methods deliver quantum chemical bond definitions, their dependence on electronic structure calculations limits their applicability in high-throughput cheminformatics. In a discipline where even SQM methods impose substantial computational demands, such approaches prove impractical for large-scale workflows. Assuming that the PubChem database contains approximately 123 million compounds,[84] and that a GFN2-xTB single-point energy calculation takes an average of 0.25 seconds per molecule, it would require approximately 356 core days to compute the connectivity of the entire database, which is impractical for a molecular identifier.

In cheminformatics, a pragmatic alternative is to infer bonding from interatomic distances.

Two well-established open-source toolkits, `RDKit`[86] and `OpenBabel`,[85] both implement distance-based bond assignment algorithms. In `OpenBabel`, a bond is created if the distance between two atoms is less than the sum of their covalent radii plus a tolerance value of 0.45 Å.[85] The values published in Cordero et al. are used for covalent radii.[144] `RDKit` uses the same approach, termed the connect-the-dots (CTD) algorithm. While CTD is the standard method, two other approaches are also available in `RDKit`.[86] The second approach is the van der Waals method, which uses the same covalent radii as the CTD method, but scales each radius by a factor of 1.3 before comparing the sum with the interatomic distance. The third approach uses an EHT calculation to determine if two atoms are bonded. The yet another extended Hückel molecular orbital package (YAeHMOP)[145] theory level is used, which is a fast EHT method implemented in `RDKit`. Using this method, the overlap population matrix is calculated, and a bond is assigned if the population exceeds a threshold of 0.15.[86]

Force fields like GFN-FF, developed by Spicher and Grimme, also require accurate connectivity information to calculate force field energy terms. GFN-FF determines this connectivity directly from the 3D coordinates of the molecule. To achieve this, GFN-FF uses environment-dependent covalent radii for each atom, which are dynamically adjusted based on the element type, partial charges, and coordination number.[146] These radii were parametrized to reproduce equilibrium bond lengths at the PBEh-3c[147] and B97-3c[148] levels of theory. The transferability of this approach has enabled its successful application across diverse systems, including organometallic complexes, supramolecular structures, and biomolecules.[149–151] The GFN-FF algorithm is summarized as follows: First, a reference covalent radius is defined for each atom based on its element type and local environment:

$$r_i^{\mathrm{ref}} = r^0(Z_i) + \Delta r_{\mathrm{shift}}(\mathrm{CN}_i, Z_i), \tag{5.1}$$

where $r^0(Z_i)$ is the base covalent radius for element Z_i, and $\Delta r_{\mathrm{shift}}(\mathrm{CN}_i, Z_i)$ is a correction based on the atomic coordination number CN_i and element Z_i. The shift is given by

$$\Delta r_{\mathrm{shift}}(\mathrm{CN}_i, Z_i) = f_{\mathrm{CN}}(Z_i)\mathrm{CN}_i, \tag{5.2}$$

where f_{CN} is a fit parameter for each element. The reference distance between two atoms i and j is then calculated as the sum of their reference covalent radii, adjusted for the electronegativities χ_i and χ_j and the adjustment parameters per period in the periodic table c_1 and c_2:

$$r_{ij}^0 = (r_i^{\mathrm{ref}} + r_j^{\mathrm{ref}})(1 - c_1|\chi_i - \chi_j| - c_2|\chi_i - \chi_j|^2) \tag{5.3}$$

The final reference distance is further modified by partial charges q_i, a charge-dependent fit parameter per element $f_q(Z_i)$ and element-specific scaling factors $f_z(Z_i)$:

$$r_{ij}^{\mathrm{ref}} = \Big(r_{ij}^0 - q_i f_q(Z_i) - q_j f_q(Z_j)\Big) f_z(Z_i) f_z(Z_j) \tag{5.4}$$

The charges are initially set to zero. Based on this reference distance, two atoms are considered

bonded if their interatomic distance r_{ij} satisfies

$$r_{ij} < f_{\text{scale}} r_{ij}^{\text{ref}}, \tag{5.5}$$

where f_{scale} is a global scaling factor (typically 1.25). Based on this initial connectivity assignment, GFN-FF introduces an iterative refinement of the molecular topology by calculating the partial charges using the electronegativity equilibration (EEQ) model, which approximates the isotropic electrostatic energy.[146] Instead of using geometric distances r_{ij}, a topological distance d_{ij} is used to calculate the EEQ charges. After updating the charges, the covalent radii and bonding assignments are recalculated. This process is usually repeated once to achieve convergence.

5.2.2 MolBar Topology Model

The absence of a universal criterion to determine molecular connectivity from 3D coordinates poses a fundamental challenge in the intersection of quantum chemistry and cheminformatics. In general, parsers can convert Cartesian coordinates into CT files which can then be used to calculate molecular identifiers like SMILES or InChI. However, different parsers, each based on different bond definitions, can generate inconsistent molecular identifiers even from identical 3D coordinates. This inconsistency highlights a critical requirement: molecular identifiers intended for quantum chemistry applications must explicitly incorporate the specific bond definition algorithm used to derive connectivity from 3D coordinates. Doing so is indispensable for maintaining reproducibility and consistency across computational analyses.

Thus, an algorithm to determine molecular connectivity from 3D coordinates is integrated into MolBar. While a previous MolBar version relied on the GFN-FF algorithm for connectivity assignment, a new method specifically tailored to MolBar was developed in this work. A deliberately simpler approach is chosen compared to the previous version, omitting environment-dependent covalent radii and iterative refinement as in GFN-FF. Now, a distance-based approach inspired by the D3 dispersion model of Grimme and coworkers[152] is adopted. This approach is similar to the van-der-Waals method implemented in `RDKit`.[86] Within D3, the floating point coordination number $\tilde{\text{CN}}_i$ for atom i is computed as:

$$\tilde{\text{CN}}_i = \sum_{j \neq i}^{N_{\text{atoms}}} \frac{1}{1 + e^{-k_1^{\text{D3}}(k_2^{\text{D3}}(r_i^{\text{cov}} + r_j^{\text{cov}})/r_{ij} - 1)}} \tag{5.6}$$

Here, r_i^{cov} and r_j^{cov} represent the covalent radii of atoms i and j, respectively, while r_{ij} denotes the interatomic distance. The parameters $k_1^{\text{D3}} = 16$ and $k_2^{\text{D3}} = 4/3$ were optimized to replicate coordination numbers derived from DFT calculations. The following criterion is used to determine the connectivity between two atoms:

$$r_{ij} < k_2^{\text{D3}}(r_i^{\text{cov}} + r_j^{\text{cov}}) \tag{5.7}$$

This expression corresponds to Equation 5.6 when the contribution of a neighbor j to the coordination number of i equals 0.5. Exploratory tests revealed that while this method performs adequately for organic molecules, it tends to underperform for transition metal complexes. To enhance accuracy for transition metal complexes, two distinct scaling factors for metal radii and for non-metal radii are introduced. Consequently, Equation 5.7 is reformulated as:

$$r_{ij} < (f_r(Z_i)r_i^{\mathrm{cov}} + f_r(Z_j)r_j^{\mathrm{cov}}) \tag{5.8}$$

In this context, f_r is a scaling factor that depends on whether the element is a metal or a non-metal. For non-metal elements, f_r is fixed at $4/3$, consistent with the original D3 model. The covalent radii of Pyykkö and Atsumi[153] are utilized for all elements, as in the original D3 framework. The metal scaling factor $f_r(Z_i)$ for each transition metal element is determined by a fitting procedure described in the next section.

5.2.3 Dataset and Computational Details

For fitting the metal scaling factor in Equation 5.8, a training and test dataset was generated from the tmQM dataset.[57] This dataset includes 108,538 transition metal complexes with geometries optimized at the GFN2-xTB[46] level of theory and Wiberg bond order (WBO) computed at the same level. For comparison of MolBar's topology assignment, GFN-FF and `RDKit`'s CTD algorithm were used to assign connectivity for the same set of molecules. Since only the metal scaling factor required fitting, only metal–nonmetal atom pairs were considered for the fitting procedure, while nonmetal–nonmetal pairs were excluded. To obtain the binary reference labels for the fitting procedure, a WBO threshold was applied. Any WBO threshold is inherently arbitrary, as there is no universal consensus on what constitutes a bond based on WBO. To identify a reasonable threshold, various values were evaluated by calculating the F1-score for connectivity assignments using GFN-FF and `RDKit`'s CTD algorithm. The threshold ($\mathrm{WBO} > 0.1695$) that maximized the F1-score for both methods was selected as the ground truth for bonding. Interatomic distances were calculated and spurious long-range contacts were filtered out by comparing interatomic distances based on the sum of covalent radii plus a 1.0 Å tolerance. Only atom pairs with interatomic distances less than this threshold were retained for further analysis. To ensure robust model evaluation, the dataset was split into training and test sets using a stratified approach. This ensured that both subsets maintained similar distributions of bonded and non-bonded pairs. The data were partitioned into 80% training and 20% test subsets. The fitting of the metal scaling factors in Equation 5.8 was performed using the training set. For each data point, the threshold is calculated as follows:

$$t = \frac{r_{ij} - f_r(Z_j)r_j^{\mathrm{cov}}}{r_i^{\mathrm{cov}}}, \tag{5.9}$$

where Z_j is the non-metal atom, $f_r(Z_j) = 4/3$, and r_j^{cov} is its covalent radius. The optimal metal scaling factor $f_r(Z_i)$ was determined by maximizing the F1-score of bond assignments against the WBO ground truth.

For comparison, the CTD algorithm as implemented in `RDKit` version 2025.9.1 was also evaluated on the hold-out test set, which is referred to as CTD/RDKit in the following. Further, GFN-FF as implemented in the `xtb` program version 6.7.1 was used to generate molecular topologies for the same set of molecules. The topology was extracted via the `-wrtopo nb` option, which outputs the neighbor list based on the GFN-FF bonding criteria.

5.2.4 Results and Discussion

The evaluation of three topology assignment methods, MolBar, GFN-FF, and CTD/RDKit, was conducted using a hold-out test set of 21,708 transition metal complexes from the tmQM dataset.[57] The fitting procedure revealed a scaling factor for metal radii of 1.15, compared to the fixed scaling factor of 1.33 for non-metal elements. The radii of transition metals are thus scaled down by approximately 13.6% compared to the original D3 model. As summarized in Table 5.1, all three methods achieve outstanding overall accuracy (99.95–99.96%) and near-perfect performance for organic bonds, with F1-scores $\geq$ 99.54%. This underscores their robust and reliable assignment of organic bond topologies. However, the dataset is highly imbalanced, as non-bonds significantly outnumber bonds. Thus, even a model predicting no bonds would still achieve high accuracy.

Table 5.1: Classification performance metrics for topology assignment methods compared against WBO ground truth calculated at GFN2-xTB level of theory. Results based on hold-out test set containing 21,708 transition metal complexes from the tmQM dataset.[57]

Method	Category	Accuracy	Precision	Recall	F1-Score
MolBar	Total	99.95	99.53	98.82	99.17
GFN-FF	Total	99.96	99.84	98.74	99.29
CTD/RDKit	Total	99.95	99.73	98.42	99.07
MolBar	Organic	99.99	99.96	99.47	99.72
GFN-FF	Organic	99.99	99.99	99.46	99.73
CTD/RDKit	Organic	99.98	100.00	99.09	99.54
MolBar	Metal	98.74	94.56	91.56	93.03
GFN-FF	Metal	98.98	98.02	90.69	94.21
CTD/RDKit	Metal	98.88	96.57	91.06	93.73

A more detailed analysis focusing on precision and recall, particularly of metal-containing bonds, reveals subtle yet important differences in their performance. High precision values indicate that when a method predicts a bond, it is highly likely to be correct. For metal bonds, GFN-FF exhibits the highest precision (98.02%), followed by CTD/RDKit (96.57%) and

MolBar (94.56%). Conversely, recall, which measures the ability to identify all actual bonds, shows an inverse trend: MolBar leads with 91.56%, followed by CTD/RDKit (91.06%) and GFN-FF (90.69%). This suggests that CTD/RDKit is conservative in predicting metal bonds, minimizing false positives but missing a few true bonds. MolBar, on the other hand, captures more true metal bonds but at the cost of increased false positives. GFN-FF strikes a balance, achieving the highest F1-score (94.21%), which harmonizes precision and recall. Despite these differences, all three methods are highly comparable with only minor differences.

The distributions of false positives and false negatives per molecule, illustrated in Figure 5.1, provide further insight into the error profiles of each method.

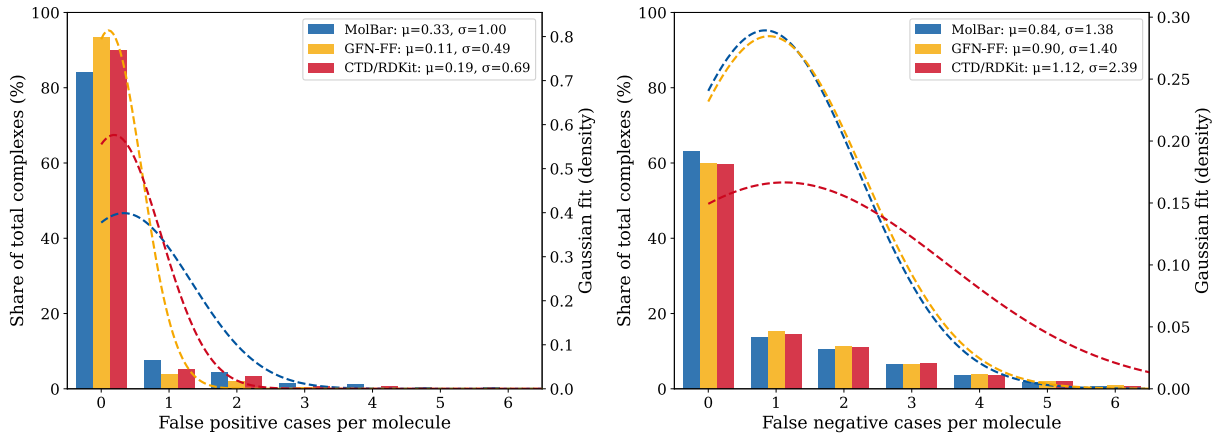
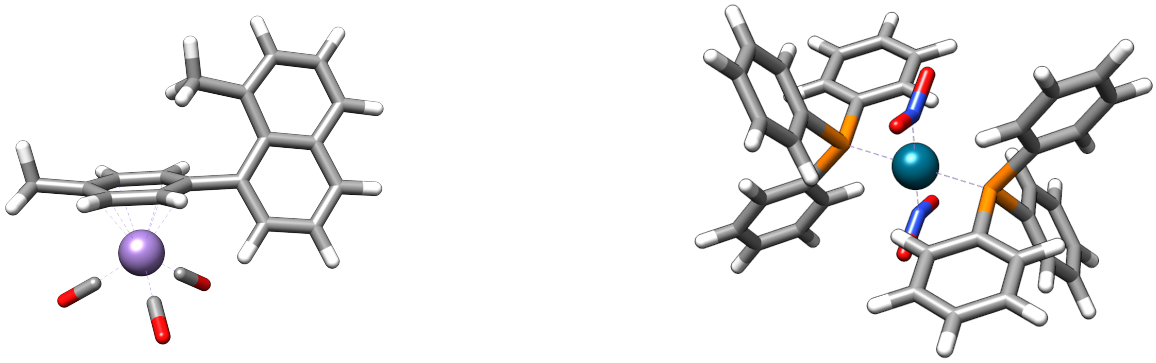


Figure 5.1: Distributions of false positive and false negative rates per molecule in the hold-out test set in the tmQM dataset.[57] Structures were optimized at GFN2-xTB[46] level of theory. MolBar and GFN-FF show tighter false negative distributions compared to CTD/RDKit, while CTD/RDKit is better at minimizing false positive cases compared to MolBar. GFN-FF achieves the best balance between minimizing false positive and false negative cases.

CTD/RDKit and GFN-FF exhibit tighter false positive distributions, with lower mean rates ($\mu_{\text{CTD/RDKit}} = 0.19$, $\mu_{\text{GFN-FF}} = 0.11$) and narrower standard deviations ($\sigma_{\text{CTD/RDKit}} = 0.69$, $\sigma_{\text{GFN-FF}} = 0.49$), indicating fewer spurious bond assignments compared to MolBar ($\mu_{\text{MolBar}} = 0.33$, $\sigma_{\text{MolBar}} = 1.00$). Nevertheless, the differences in false positive rates are minor. For false negatives, MolBar and GFN-FF demonstrate more consistent performance compared to CTD/RDKit, with lower mean rates ($\mu_{\text{MolBar}} = 0.84$, $\mu_{\text{GFN-FF}} = 0.90$) and reduced variance ($\sigma_{\text{MolBar}} = 1.38$, $\sigma_{\text{GFN-FF}} = 1.40$), suggesting they are less likely to miss existing bonds compared to CTD/RDKit ($\mu_{\text{CTD/RDKit}} = 1.12$, $\sigma_{\text{CTD/RDKit}} = 2.39$). While MolBar and GFN-FF are comparable in false negative performance, CTD/RDKit shows a broader distribution, indicating more cases with higher false negative rates. Overall, the differences between all three methods are marginal, underscoring their comparable performance in topology assignment for transition metal complexes.

At first glance, the prevalence of false negative outliers is seemingly substantial across all three methods, especially when contrasted with the relatively low false positive rates. However, a closer analysis reveals inherent challenges in relying on WBO for bonding assignment

(Figure 5.2). For instance, in the ABITEN complex, all three methods correctly refrain from assigning bonds between the Mn center and the oxygen atoms of the CO ligands. Similarly, in the ACOWIA complex, none of the methods classify the interactions between the Pd center and the oxygen atoms of the NO_2 ligands as bonds. Despite WBO values of around 0.3, suggesting potential bonding, chemical reasoning confirms that no bonds should exist in these cases, as coordination to the metal center occurs via carbon or nitrogen atoms, rather than oxygen. The non-zero WBO values observed between the metal and ligand oxygen atoms are likely due to π-backbonding effects. These examples underscore the limitations of WBO as a bonding descriptor and the complexities of accurately characterizing bonding in such systems. Not all outliers stem from such ambiguous cases, however. In some rare cases, MolBar misses the η^6-benzene ligand in ABITEN, for example.



(a) ABITEN complex **(b)** ACOWIA complex

Figure 5.2: Analysis of bonding outliers in transition metal complexes from the tmQM dataset.[57] (a) In ABITEN, all three methods correctly do not assign Mn–O bonds, despite WBO ≈ 0.3 suggesting otherwise. MolBar misses the η^6-benzene ligand as a rare outlier. (b) In ACOWIA, all three methods correctly do not assign Pd–O bonds, despite significant WBO ≈ 0.3 values. These cases illustrate the limitations of WBO as a bonding descriptor and the challenges in interpreting bonding interactions using distance-based topology assignment methods. Such cases most likely arise from π-backbonding effects rather than true bonding interactions.

In summary, the three methods exhibit remarkably similar performance when inferring molecular connectivity from 3D coordinates, with no single approach demonstrating a clear advantage. When applied to the assignment of organic bonds, all three methods prove to be both highly accurate and reliable, showcasing robust capabilities across diverse molecular scenarios. For metal bonds, however, their strengths diverge: MolBar prioritizes recall, GFN-FF adopts a balanced approach, and CTD/RDKit emphasizes precision. This nuanced distinction suggests that the optimal method depends on the specific demands of the application at hand. The marginal variations in their F1-scores further underscore the reliability and resilience of each approach. GFN-FF stands out among them by achieving a balance between precision and recall. This is likely due to its iterative refinement process for topology assignment. Yet, this sophistication introduces a trade-off in complexity, as GFN-FF needs to solve the EEQ eigenvalue problems with multiple iterations. By relying solely on geometrically derived covalent radii, MolBar and CTD/RDKit provide a computationally simple alternative that sacrifices lit-

tle performance. Given their comparable accuracy, the proposed MolBar algorithm qualifies as a practical solution, offering a straightforward and efficient method for determining molecular connectivity from 3D coordinates in quantum chemistry applications.

5.3 Rigidity Matrix: Electronic Hindrance Classifier

5.3.1 Introduction

The set of electronically hindered bonds, denoted by $M_{\text{electronic}}$, is necessary for the computation of the rigidity matrix $\mathcal{I}$ (Equation 4.7).[154] In conjunction with the set of rings C^s and sterically hindered bonds M_{steric}, this set enables the determination of substructures in molecules that are characterized by limited degrees of conformational freedom. With these fragments, the topography of a molecule can be described by Equation 4.4 without including conformational information. $M_{\text{electronic}}$ includes bonds of elevated bond order, such as aromatic, double, or triple bonds. As the overlap of the π-orbitals would become reduced in the transition state, these bonds can show elevated rotational barriers. Therefore, rotation around such bonds is not indicative of conformational changes; rather, it results in a change in the identity of the molecule. While C-C double bonds present rotational barriers of approximately 60 kcal/mol, C-C single bonds permit free rotation with a rotational barrier of only about 3 kcal/mol. Intermediate bonds, including aromatic bonds or bonds with partial double-bond character, have been observed to contribute to an increase in rotational rigidity, even in the absence of close steric contacts. A classical example is the restricted rotation around the amide C–N bond, which is attributed to resonance stabilization and leads to barriers of 15–25 kcal/mol, even in small, sterically unhindered systems.[155] Thioamides manifest elevated barriers, ranging from 20 to 30 kcal/mol, although being formally a single bond.[120,156]

In summary, in order to obtain the desired set $M_{\text{electronic}}$, it is necessary to consult how bond order information could be implemented into MolBar. In the domain of quantum chemistry, a range of methodologies exists for the calculation of bond order through electronic structure calculations. As delineated in Section 5.2, such methods have been discussed for the purpose of ascertaining connectivity. Using WBOs, a value above one indicates an increased bond order, suggesting the presence of a (partial) double bond or triple bond.[139] Such indices in MolBar would provide the desired results once a threshold value is set to separate electronically hindered from unhindered bonds. Nevertheless, as previously mentioned, any electronic structure calculation is computationally too expensive in the context of an identifier like MolBar. One example from cheminformatics that is worth mentioning, however, is the generation of a metal complex graph representation with bond order information from natural bond orbitals (NBO) calculations, starting with 3D coordinates. This process is explained by Kneiding et al. in their paper.[157]

In the domain of cheminformatics, a range of methodologies have been developed to derive bond orders without the necessity of conducting electronic structure calculations. These methodologies employ elemental and connectivity information, in some cases in conjunction with select geometric data, to facilitate the estimation of bond orders. In the widely used cheminformatics program `OpenBabel`, atoms are assigned their hybridization state based on bond angles and geometries.[85] Atoms marked as unsaturated are checked for unsaturated neighbors

to assign double or triple bonds. Similar work has been published by Baber and Hodgkin,[158] Hendlich et al.,[159] Sayle[160] and Labute.[161] In a more recent development, Y. Kim and W. Kim unveiled an algorithm that operates independently of geometry information.[162] This algorithm relies solely on connectivity and element-specific valences, disregarding geometry-related parameters. Two distinct matrices are employed in this analysis: the adjacency matrix a_{ij}, which describes the connectivity between atoms, and the bond order matrix BO_{ij}, which includes the integer bond orders for a given bond (i, j) in the adjacency matrix. The unsaturation degree u_i can be calculated based on the valences $N_{v,i}$ for a given atom, contingent upon its element type, and the adjacency matrix a_{ij}:

$$u_i = N_{v,i} - \sum_{j \neq i}^{N_{\text{atoms}}} a_{ij} \tag{5.10}$$

It is worth noting that a variety of bond order matrices can be derived from a specific adjacency matrix. However, by employing a set of explicit criteria, it is possible to identify an optimal bond order matrix that fulfills two conditions. First, the element-wise difference between the adjacency matrix and the bond order matrix must yield the total degree of unsaturation:

$$\sum_{ij} (\mathrm{BO}_{ij} - a_{ij}) = \sum_{i} u_i \tag{5.11}$$

This information delineates the presence or absence of unsaturated atoms. Further, charge conservation with the total charge of the molecule q_{mol} must be fulfilled:

$$\sum_{i} q_i = q_{\text{mol}} \tag{5.12}$$

The partial charges, denoted by q_i, can be derived from tabulated charge values based on the valence of an atom. In the event that multiple bond order matrices meet both criteria, the one with the higher sum of all matrix elements $\sum_{ij} \mathrm{BO}_{ij}$ is selected, as it is associated with higher stability. The present algorithm has been implemented in the program known as `xyz2mol`, developed by Jan Jensen and co-workers.[136] This implementation has subsequently been incorporated into the de-facto standard cheminformatics library, `RDKit`.[86] However, an algorithm based on valences only functions adequately with organic molecules, where valences are clearly defined. Hence, this algorithm cannot be applied straightforwardly to metal complexes. Recently, an extension of `xyz2mol` to generate SMILES codes from 3D coordinates of metal complexes has been published by Jan Jensen and co-workers.[122] Here, the `xyz2mol` algorithm is carried out for each ligand separately, in combination with an iterative EHT calculation at YAeHMOP[145] level of theory, which is implemented in `RDKit`. The metal-ligand bond is treated as a dative bond, with the ligand donating electrons to the metal center. The metal charge is derived from the ligand charges and the total charge of the molecule, ensuring charge conservation. This iterative procedure is necessary to determine the optimal charge and bond orders for each ligand. In order to assign the property "rotatable" or "non-rotatable" in

MolBar, it is insufficient to utilize integer bond orders as provided by the cheminformatics algorithms that have been discussed. The underlying rationale for this phenomenon is that the (thio)amide C–N bond, being non-rotatable in some cases, would be depicted as a single bond under these methodological frameworks.

In recent years, a novel approach has emerged within the field of chemistry, precipitated by the advent of ML. The utilization of ML in chemistry can be traced back to the 1960s, when it was employed in the development of early QSAR models. However, recent advancements in algorithms, computational capabilities, and the substantial growth in the availability of chemical data have led to a paradigm shift in its application and impact.[163] At its core, ML models seek to learn a mapping function f that makes predictions in response to input features x:[164]

$$y = f(x, W) \tag{5.13}$$

y denotes the predicted target property, which can be reaction yields, molecular energies, or material properties. W represents the learned parameters. It is noteworthy that both x and y can represent a wide range of data types, from scalars and vectors to matrices and graphs. The learning process entails the minimization of a loss function, denoted by $\mathcal{L}$, which quantifies the discrepancy between predictions y and ground truth values $\hat{y}$:[164]

$$W^* = \underset{W}{\operatorname{argmin}} \sum_{i=1}^{n_{\text{data}}} \mathcal{L}(\hat{y}_i, y_i) \tag{5.14}$$

where n_{data} is the number of training samples. For regression tasks, common loss functions include the mean squared error (MSE), while classification tasks often utilize cross-entropy loss.[165] The selection of model architecture f is contingent upon the complexity of the problem and the underlying structure of the data. Linear regression, the simplest approach, models the relationship as follows:[166]

$$y = xW^T + b \tag{5.15}$$

where b is a bias term. Beyond linear regression, various advanced methods have been developed to capture complex, nonlinear relationships in data. Probabilistic approaches, such as Gaussian process regression (GPR), offer predictions and uncertainty estimates, which are useful for experimental validation and decision-making under uncertainty.[167] Ensemble methods, such as random forests, combine multiple weak learners to enhance robustness and predictive performance, particularly in scenarios with limited data or high noise.[168] In recent years, deep learning has made significant strides in the field of ML, including its application to chemical problems. After their initial development in the mid-20th century, neural networks experienced a resurgence in popularity during the 1980s and 1990s. This was driven by breakthroughs in computing power and the increasing availability of large datasets, which enabled the training of deeper and more complex models.[169] A development during this period was the introduction of the backpropagation algorithm by Rumelhart et al., which provided an efficient way to train neural networks by iteratively adjusting weights based on gradient descent.[170] The

advent of automatic differentiation (autograd) further revolutionized the field by automating the computation of gradients.[171] To address training instability in deep networks, normalization techniques such as BatchNorm[172] and LayerNorm[173] were introduced, which accelerated convergence and improved model performance.

At the heart of neural networks is the perceptron, a basic unit that calculates a weighted sum of inputs followed by a nonlinear activation σ:

$$h = \sigma(xW^T + b) \tag{5.16}$$

Non-linear activation functions include rectified linear unit (ReLU), sigmoid, and tanh, which enable the network to learn complex patterns and relationships in data.[174] Multi-layer perceptrons extend this concept by stacking multiple layers, with each layer transforming the representation of the previous layer.

$$h^{(l)} = \sigma(h^{l-1}(W^l)^T + b^l) \tag{5.17}$$

where l indexes the layer. Modern architectures, however, have evolved beyond dense layers by leveraging the inherent structure of the data. In the domain of molecular modeling, the graph neural network (GNN) architecture has emerged as a prominent and influential development. These networks operate directly on the intrinsic graph structure of molecules, when being described as bonded atoms. As indicated by Reiser et al., the GNN architecture is capable of learning chemically meaningful representations through the iterative aggregation of information from neighboring atoms and bonds.[175] Thus, a diverse array of neural architectures has been proposed for molecular property prediction, building on the existing momentum. These architectures encompass a range of approaches, including message-passing methods[63,176–179] and equivariant models that encode physical symmetries by design.[180–182] Many of these models fall under the general framework of message passing neural network (MPNN), as introduced by Gilmer et al.[183] Notable examples of MPNN-based models include the Directional Message Passing Neural Network[176,177] or directed message passing neural network (DMPNN) in Chemprop.[180,181] On the other hand, architectures such as SchNet,[63] physical neural network (PhysNet),[178] hierarchically interacting particle neural network (HIP-NN)[179,184] or equivariant graph neural network (EGNN)[65,182] deviate from the MPNN formalism, although they share to some extent the same underlying principles.

MPNNs operate on molecular graphs $G = (V, E)$. The MPNN iteratively refines the feature representations of atoms and bonds through a message-passing process, in which each atom aggregates messages from its neighbors to update its representation. The message-passing step is typically defined as:

$$m_i^l = \sum_{j \in N(i)} M^l(h_i^{l-1}, h_j^{l-1}, e_{ij}), \tag{5.18}$$

where m_i^l is the message for atom i at layer l, $N(i)$ denotes the neighbors of atom i, and M^l

is a message function that combines the features of the source atom i, the destination atom j, and the edge feature e_{ij}. Here, $h_i^{l-1} \in \mathbb{R}^{n_{\text{atoms}} \times \mathcal{H}}$ represents the feature vector of atom i at layer $l-1$, and $e_{ij} \in \mathbb{R}^{n_{\text{bonds}} \times \mathcal{H}}$ represents the bond feature between atoms i and j. $\mathcal{H}$ is the hidden dimension, while n_{atoms} and n_{bonds} denote the number of atoms and bonds in the molecule, respectively. Using the aggregated messages, the atom features are updated as:

$$h_i^l = U^l(h_i^{l-1}, m_i^l), \tag{5.19}$$

where U^l is an update function combining the previous atom features with the aggregated messages. After L message-passing layers, the final atom representations can be used to predict molecular properties. For example, the final atom features can be aggregated into a global molecular representation:

$$y = R(h_i^L | i \in V), \tag{5.20}$$

where R is a readout function that maps the set of atom features into a single vector y representing the molecule. This molecular representation can then be used to predict properties such as energies, forces, or other molecular descriptors. For predicting bond orders, the MPNN architecture is well-suited, as it can effectively capture the local chemical environment around each bond. Architectures like DMPNN[180] and the HIP-NN[184] have shown good performance in prior studies for bond order prediction tasks. In particular, DMPNN has been applied in the work of Li et al.,[185] whereas HIP-NN has been used by Magedov et al.[186]

5.3.2 Electronic Hindrance Model

As outlined in Chapter 3, a bond is classified as electronically hindered when its rotational barrier surpasses 20 kcal/mol, assuming no steric effects are present. The following Section 5.4.2 introduces a computational workflow to calculate rotational barriers at the UMA[65]//GFN2-xTB[46] level of theory. More details about this dataset and workflow can be found in the following section. The rotational barriers as a function of the WBO is shown in Figure 5.3. High rotational barriers above 20 kcal/mol with WBO close to one can be attributed to steric effects. For this section, the rotational barriers with high WBO values are of particular interest, as they indicate the presence of electronic hindrance. The analysis of this dataset revealed that a threshold of 1.25 corresponds to a rotational barrier of roughly 20 kcal/mol.

In principle, any semiempirical or DFT method could be employed to calculate WBOs. However, these methods are too computationally demanding for their intended use in MolBar. Thus, a ML model is chosen to predict whether a bond is electronically hindered based on its WBO from a reference method. The MPNN architecture was chosen as it operates directly on molecular graphs, making it inherently insensitive to small fluctuations in atomic positions. Once the molecular graph is constructed in Section 5.2, the classification is always identical for different conformations of the same molecule as it is not influenced by varying dihedral angles. A neural network approach was chosen over other rule-based approaches due to the

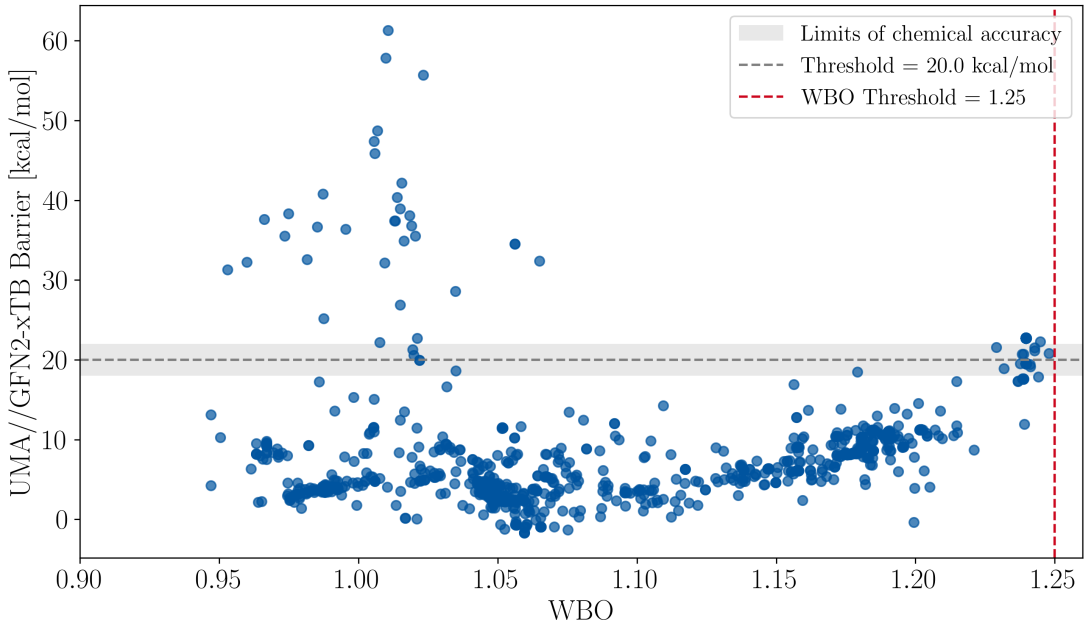


Figure 5.3: Wiberg bond order (WBO) plotted against rotational barrier for the dataset discussed in Section 5.4.2, calculated at the UMA[65]//GFN2-xTB[46] level of theory. The horizontal dashed line at 20 kcal/mol marks the threshold applied in MolBar, while the vertical dashed line at WBO = 1.25 indicates a proposed cutoff for electronic hindrance in the absence of steric effects. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

availability of large datasets and the ability to more easily generalize towards transition metal complexes. This approach also allows for the construction of a computationally lightweight model, as it only requires a few matrix multiplications per molecule. For the exact number of operations, see the end of this section. This makes the approach suitable for integration into the MolBar workflow.

Exploring different architectures led to the conclusion to remain close to the original MPNN formulation,[183] instead of using DMPNN or HIP-NN. Each atom is represented by a feature vector $x_i \in \mathbb{R}^{n_{\text{atoms}} \times F_{\text{atom}}}$, where F_{atom} denotes the atom feature dimension. Similarly, each bond is characterized by a feature vector $e_{ij} \in \mathbb{R}^{n_{\text{bonds}} \times F_{\text{bond}}}$, where F_{bond} is the bond feature dimension. The task is to classify each bond (i, j) by predicting the probability p_{ij}^{WBO} that its WBO exceeds a threshold of 1.25, thereby identifying it as electronically hindered.

Modeling the entire periodic table is challenging due to the vast diversity of bonding environments and electronic structures. However, rotational barriers around a bond are primarily governed by π-bonding interactions arising from p-orbital overlap. Thus, the "double bond rule" can be applied.[187] This empirical observation from inorganic and organometallic chemistry states that elements in period 3 and beyond rarely form multiple (double or triple) bonds, either with themselves or with other elements. This is because the atomic orbitals of these heavier elements (e.g., 4p) are larger and more diffuse, leading to weaker π-overlap compared to the compact 2p orbitals of carbon, nitrogen, and oxygen. Therefore, only bonds contain-

ing a subset of elements such as boron, carbon, nitrogen, oxygen, sulfur, but also silicon and phosphorus are considered, while others such as halogens, hydrogen, and transition metals are excluded. Although exceptions exist, for simplicity, bonds involving the latter elements are masked and treated as non-hindered bonds.[188]

For computational efficiency and generalizability, the feature vectors were deliberately designed to use only graph properties and tabulated values that are available for the entire periodic table. For the atom features, one-hot encoded representations of the relevant main group elements (boron, carbon, nitrogen, oxygen, silicon, phosphorus, and sulfur) are used together with their coordination numbers. For bond features, only the Pauling electronegativity difference between the two bonded atoms was considered. This design results in $F_{\text{node}} = 30$ and $F_{\text{bond}} = 1$, providing a compact representation that avoids embeddings for each atom type. Initially, atom and bond embedding layers transform these input features into a higher-dimensional space suitable for learning. This transformation involves a linear mapping followed by batch normalization and a ReLU activation:

$$h_i^{(0)} = \text{ReLU}(\text{BatchNorm}(W_{\text{node}} x_i)), \quad W_{\text{node}} \in \mathbb{R}^{\mathcal{H} \times F_{\text{node}}}, \tag{5.21}$$

where W_{node} is a weight matrix that maps the input atom features to the hidden dimension $\mathcal{H}$. Similarly, bond features are transformed to the same hidden dimension $\mathcal{H}$:

$$e_{ij}^{(0)} = \text{ReLU}(\text{BatchNorm}(W_{\text{bond}} e_{ij})), \quad W_{\text{bond}} \in \mathbb{R}^{\mathcal{H} \times F_{\text{bond}}} \tag{5.22}$$

W_{bond} is again the weight matrix. Using both atom and bond embeddings, a message-passing layer is constructed that aggregates information from neighboring atoms and bonds in a directed manner. This message passing is performed for L layers, where each layer updates the atom embeddings. For each bond (i, j) and layer l, the following steps are executed: First, a message is computed using the source atom $h_i^{(l-1)}$, the destination atom $h_j^{(l-1)}$, and the bond feature e_{ij}:

$$m_{ij} = \text{MLP}([h_i^{(l-1)} \| h_j^{(l-1)} \| e_{ij}]), \quad m_{ij} \in \mathbb{R}^{\mathcal{H}} \tag{5.23}$$

Here, $\|$ denotes concatenation, and MLP is a multilayer perceptron defined as:

$$\begin{aligned} &\text{MLP}([h_i^{(l-1)} \| h_j^{(l-1)} \| e_{ij}]) = \\ &W_2 \cdot \text{ReLU}(\text{BatchNorm}(W_1 [h_i^{(l-1)} \| h_j^{(l-1)} \| e_{ij}])) \\ &\quad W_1 \in \mathbb{R}^{\mathcal{H} \times (3\mathcal{H})}, \quad W_2 \in \mathbb{R}^{\mathcal{H} \times \mathcal{H}} \end{aligned} \tag{5.24}$$

where W_1 and W_2 are weight matrices, and BatchNorm is applied to stabilize the intermediate activations. For each atom j, messages from all connected bonds that lead to atom j are aggregated:

$$h_j^{(l)} = \text{ReLU}(\text{BatchNorm}(\sum_{i \in \mathcal{N}(j)} m_{ij})) \tag{5.25}$$

After L layers of message passing, the final atom embeddings are used to predict bond properties. Unlike the original MPNN formulation, no readout function to aggregate atom embeddings into a single graph-level representation is required for this task. Instead, bond properties are directly computed from the final atom embeddings. For each bond (i, j), the embeddings of the source and destination atoms are concatenated and passed through an MLP:

$$y_{ij}^{\text{WBO}} = \text{MLP}([h_i^{(L)} \| h_j^{(L)}]), \quad y_{ij}^{\text{WBO}} \in \mathbb{R} \tag{5.26}$$

The model outputs logits y_{ij}^{WBO} for each bond. Since these logits are directed, the logits over both directions are averaged to obtain a single score for each bond:

$$\bar{y}_{ij}^{\text{WBO}} = \frac{y_{ij}^{\text{WBO}} + y_{ji}^{\text{WBO}}}{2} \tag{5.27}$$

The probability that the bond (i, j) is electronically hindered is then computed using a sigmoid function:

$$p_{ij}^{\text{WBO}} = \frac{1}{1 + e^{-\bar{y}_{ij}^{\text{WBO}}}} \tag{5.28}$$

Using a defined threshold ω_{WBO}, the model can then classify the bond (i, j) as electronically hindered or not, leading to the set of electronically hindered bonds:

$$M_{\text{electronic}} = \{(i, j) \mid \forall (i, j) \in E(G) \wedge p_{ij}^{\text{WBO}} > \omega_{\text{WBO}}\} \tag{5.29}$$

MolBar incorporates an additional geometric criterion: if the dihedral angle between the two atoms adjacent to a bond exceeds 45 degrees, the bond is deemed electronically unhindered, since non-planar configurations generally do not facilitate substantial π-conjugation. The model architecture is designed to remain computationally efficient, requiring only $2 + 2L$ matrix multiplications per molecule. This efficiency arises from the use of two embedding layers followed by two linear transformations per message-passing layer. Empirical testing determined that five message-passing layers deliver optimal performance, leading to a total of 12 matrix multiplications per molecule.

5.3.3 Dataset

Developing a robust WBO classifier requires a comprehensive dataset that captures both the chemical diversity of organic molecules and the electronic complexity near challenging environments, such as transition metal centers. While the model explicitly excludes transition metal bonds through masking, it must still be trained on data that includes such bonds to ensure accurate predictions for main-group bonds in the vicinity of metals and to enhance generalization to organometallic systems. To meet these criteria, the QMugs dataset is used, which provides high-quality semi-empirical quantum chemical data for over 665,000 molecules derived from the ChEMBL database, spanning approximately two million conformers.[56] All

properties and geometries were computed at the GFN2-xTB[46] level of theory. To broaden chemical coverage to metal-containing systems, additionally the tmQM dataset[57] is incorporated, which comprises over 108,000 transition metal complexes, covering 30 metals in the 3d, 4d, and 5d series, and over 30,000 unique ligands. The 2024 release includes optimized geometries, natural atomic charges, and WBOs, capturing a diverse range of bonding motifs around metal centers. QMugs and tmQM are combined to construct a training set with electronically diverse environments. Bonds are identified consistently across datasets using MolBar's topology function from Section 5.2, applied to 3D molecular coordinates, with WBO annotations extracted from the respective datasets.[189] The full dataset comprises 774,450 molecules and over 46 million bonds, each annotated with WBO values. For each molecule, atom features were computed as detailed in Section 5.3.2, and molecular graphs were stored as `PyTorch Geometric` data objects, containing graph connectivity, atom features, and bond-level WBO targets.

5.3.4 Model Training and Evaluation

The electronic hindrance classification model was trained using `PyTorch`[190] version 2.6.0 and `SciPy`[135] version 1.14.1. Additionally, `PyTorch Geometric`[191] version 2.6.1 was utilized.

Parameter optimization relied on the Adam optimizer, which was initialized with a learning rate of 10^{-3}. To adapt the step size dynamically, a learning rate scheduler was incorporated whenever the validation loss reached a plateau. The cross-entropy loss function served as the objective for model optimization. Regularization was enforced through early stopping, halting training if validation loss failed to improve over five successive epochs. The model weights associated with the lowest validation loss were preserved for the final assessment.

Training and evaluation followed a five-fold CV approach. The avoidance of data leaks between training and test datasets was taken into account, as such leaks carry the risk of overstating performance metrics and undermining generalizability. Following recommendations from [192], various data-splitting methods were assessed, including random, scaffold-based,[193] and uniform manifold approximation and projection (UMAP)-based strategies.[194]

Random Split

In the random split strategy, molecules are assigned to training and test sets without considering structural relationships. While straightforward to implement, this approach may result in structurally similar compounds appearing in both sets, potentially leading to overly optimistic performance estimates.

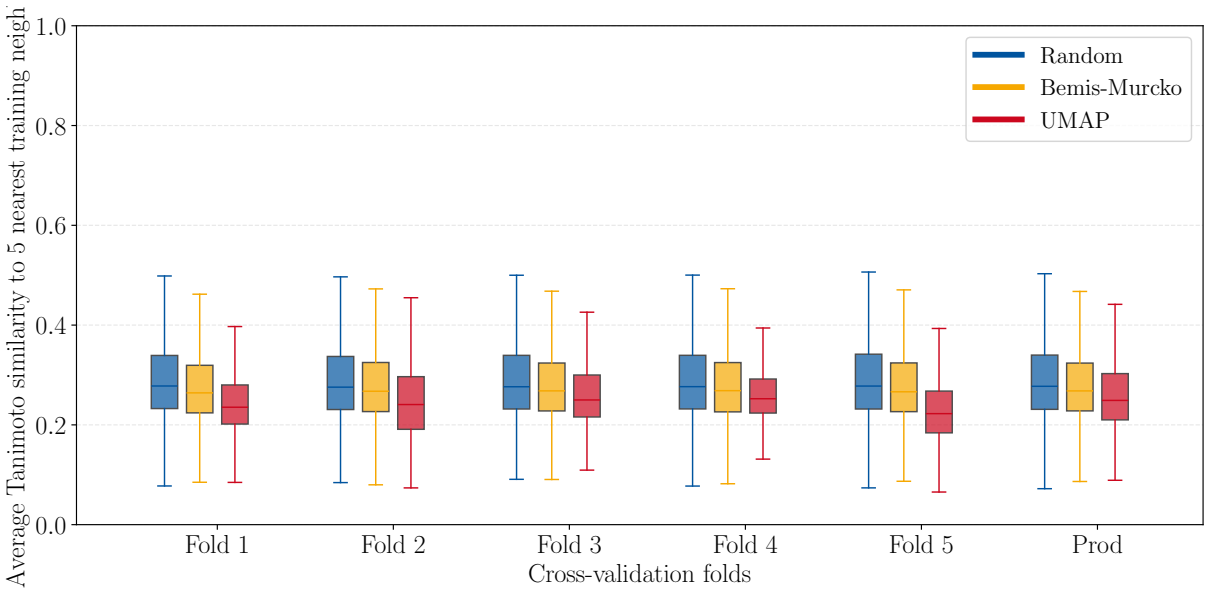


Figure 5.4: Comparison of data-splitting methods based on average Tanimoto similarity between test fragments and their five nearest neighbors in the training set. The UMAP-based and Butina-based methods demonstrate the lowest similarity, indicating superior structural dissimilarity between training and test sets. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

Scaffold Split

In the scaffold-based split, molecules are grouped by their Bemis–Murcko scaffolds using `RDKit`.[86] All molecules sharing the same scaffold are assigned exclusively to either the training or the test set.[193] This strategy enforces a separation based on core molecular frameworks and promotes evaluation on structurally distinct compounds.

UMAP-based Split

In the UMAP-based splitting strategy, Morgan fingerprints were embedded into a two-dimensional space using the UMAP algorithm with parameters `n_neighbors = 7` and `min_dist = 0.1`. The Python implementation provided by the `umap-learn` package (version 0.5.3) was employed.[195] Subsequently, the embedded representations were clustered using the KMeans algorithm with `n_clusters = 9`, as implemented in `scikit-learn` (version 1.8.0).

Figure 5.4 illustrates the similarity of the test set compared to the training set for each fold and each splitting method. Each splitting method was implemented within the five-fold CV framework. For every fold, structural similarity between training and test sets was measured by calculating the average Tanimoto similarity using Morgan fingerprints. This involved comparing each test fragment's fingerprint to its five closest neighbors in the corresponding training set.

The UMAP-based partitioning was selected due to its ability to enforce structural dissimilarity between training and test sets more effectively than random and scaffold-based methods.

With this UMAP-based partitioning, all data were split into the data used for CV and a hold-out test set to monitor the production model's generalization (80/20 split). Then, the cross-validation (CV) data was reshuffled and partitioned into five folds, where each fold consists of a training, validation, and test set using the UMAP-based method (64/16/20 split within each fold). For all models, the threshold for classifying bonds as electronically hindered was determined by maximizing the F1-score on the validation sets. Performance metrics—accuracy, precision, recall, F1-score, and ROC-AUC—were computed independently for each fold. Mean values and standard deviations of these metrics across the five folds were used to assess both predictive performance and robustness to data partitioning. Learning curves and supplementary performance visualizations can be found in the Appendix A.1.

For the comparison with `xyz2mol`, the scripts were obtained from the GitHub repository[196] and adapted to process the molecules in the test set. The `xyz2mol` implementation was executed using `RDKit` version 2024.03.3. Timings were measured as wall-clock time on an Apple M1 Pro CPU with 8 cores and 32 GB RAM, running macOS 26.1. All experiments were conducted using single-core execution, and the reported times represent the average of 5 independent runs to ensure reproducibility.

5.3.5 Results and Discussion

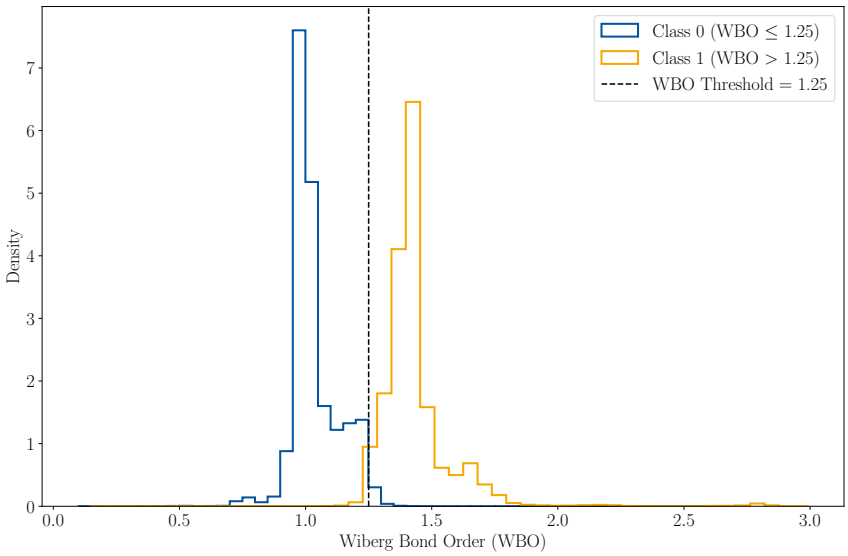
The performance metrics, averaged from five-fold cross-validation, alongside the corresponding results for the hold-out test set, are presented in Table 5.2. The trained MolBar classifier effectively distinguishes between electronically hindered and flexible bonds, demonstrating high predictive accuracy and robustness across all cross-validation folds. This consistency extends to the production model when evaluated on the independent test set.

Table 5.2: Five-fold cross-validation and hold-out test set performance of the electronic hindrance classifier. For cross-validation, metrics are averaged across the five folds, and mean values together with standard deviations are reported.

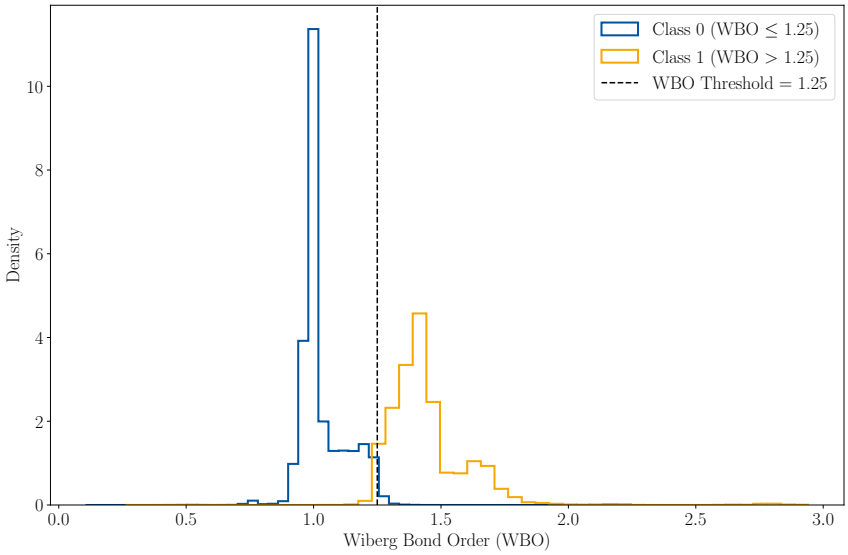
Metric	Averaged K-Fold CV	Hold-Out Test Set
Accuracy	$98.98 \pm 0.12\%$	98.30%
ROC-AUC	$99.96 \pm 0.01\%$	99.89%
Precision	$98.8 \pm 0.4\%$	98.00%
Recall	$98.8 \pm 0.4\%$	98.15%
F1-score	$98.8 \pm 0.4\%$	98.08%
Specificity	$99.05 \pm 0.15\%$	98.42%

As detailed in Table 5.2, the electronic hindrance classifier delivers consistently strong and stable performance. During five-fold cross-validation, the model attains a mean accuracy of 98.98% with a small standard deviation of $\pm 0.12\%$, indicating that predictive performance remains largely unaffected by data partitioning. The near-perfect mean ROC-AUC of $99.96 \pm 0.01\%$ confirms the classifier's robust ability to differentiate electronically hindered bonds from flexible ones across all decision thresholds, with the narrow uncertainty further emphasizing model stability. Both precision ($98.8 \pm 0.4\%$) and recall ($98.8 \pm 0.4\%$) are high and balanced, yielding an F1-score of $98.8 \pm 0.4\%$. This balance signifies that the classifier maintains sensitivity without compromising precision, rarely misclassifying either bond type. The mean specificity of $99.05 \pm 0.15\%$ further demonstrates a strong capability to correctly identify non-hindered (flexible) bonds. Collectively, the tightly grouped cross-validation metrics reflect a robust model with minimal variance. Crucially, the production model's evaluation on the independent hold-out test set reproduces strong performance across all metrics, reinforcing the model's generalizability.

Figure 5.5 depicts the distributions of predicted WBO values for the training and validation sets, as well as those of the independent test set. The classifier successfully segregates electronically hindered bonds (orange) from flexible bonds (blue), with most hindered bonds displaying reference WBO values above 1.25, while non-hindered bonds predominantly fall below this threshold. A minor overlap near the 1.25 cutoff is observed, though it remains negligible compared to the overall separation between the two classes. These findings underscore the classifier's proficiency in distinguishing between single, partial double, and full double bond character with high fidelity.



(a) Training set model performance.



(b) Test set model performance.

Figure 5.5: Distribution of the WBO grouped by the classes predicted by the final production electronic hindrance classifier for the training set (left) and the independent test set (right). The x-axis shows the ground-truth reference WBO computed at the GFN2-xTB level of theory. The histograms illustrate the density of these reference values for bonds predicted by the model as electronically hindered (Class 1, target WBO $>$ 1.25, shown in orange) and non-hindered (Class 0, target WBO $\leq$ 1.25, shown in blue). The clear separation demonstrates that the classifier effectively categorizes the bonds. Most bonds predicted as electronically hindered indeed possess reference WBOs above 1.25, while non-hindered predictions correspond to WBOs below 1.25. The small overlap near the threshold of 1.25 (dashed line) reflects both the model's predictive uncertainty and the conformational dependence of the WBO in the reference data, as the model relies solely on 2D graph representations without 3D geometric input. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

To evaluate the practical usefulness of MolBar's electronic hindrance classifier, a comparative analysis was performed with the `xyz2mol` implementation of Jan Jensen's group, which provides rule-based bond order assignments. In particular, the extended version that can handle transition metal complexes was used.[122] This method is termed xyz2mol/RDKit in the following, as it relies on `RDKit`. `xyz2mol` without transition metal complex extension already has been implemented in `RDKit`.[86] The transition metal extension will be included in an upcoming `RDKit` release, but is currently only available as a separate implementation on GitHub.[196] As described in the introduction to this chapter, the xyz2mol/RDKit algorithm assigns discrete bond types (single, aromatic, double, and triple) based on chemical heuristics and EHT calculations. For this comparison, a binary classification was derived by treating single bonds as rotatable and grouping aromatic, double, and triple bonds as non-rotatable. To account for resonances, the maximum bond order across all possible mesomeric structures was selected for each bond.

The evaluation focused on 100 randomly selected transition metal complexes from the independent hold-out test set. This subset size was chosen due to limitations in the computational stability of the xyz2mol/RDKit implementation. The absence of purely organic molecules in this subset presents a more difficult scenario, as the electronic environment of metal complexes is significantly more challenging.

Figure 5.6 compares the distribution of reference WBO values stratified by predicted bond classes from MolBar's MPNN classifier and xyz2mol/RDKit. MolBar's MPNN classifier demonstrates an advantage over the xyz2mol/RDKit method in distinguishing rotatable from non-rotatable bonds within this challenging subset of transition metal complexes. Compared to the results in Figure 5.5, the separation achieved with MolBar shows comparable performance, albeit with a slight decline, e.g. 3% in accuracy, due to the inherent complexity of the subset under consideration. In contrast, xyz2mol/RDKit shows a tendency to overestimate WBO values, resulting in some bonds being classified as non-rotatable (orange) even though their true WBO is below 1.25. This finding is also supported by the evaluation metrics summarized in Table 5.3.

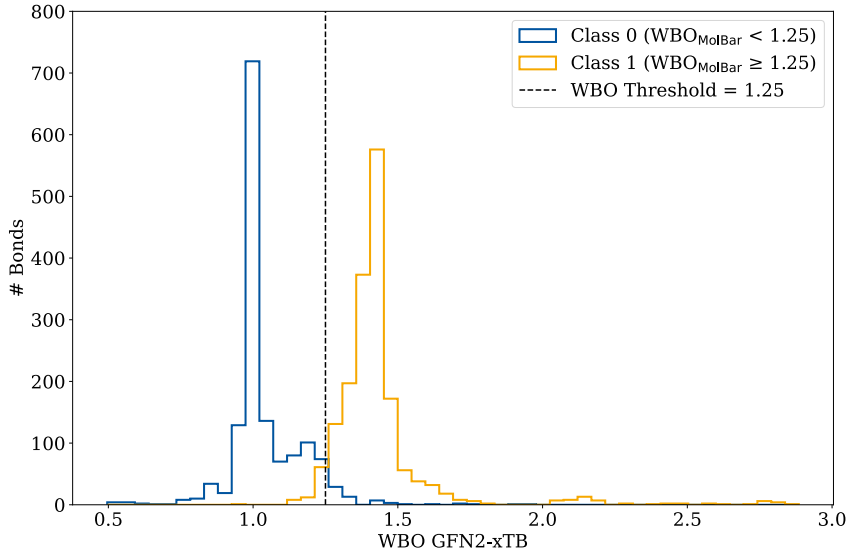
MolBar achieves an accuracy of 95.81%, outperforming xyz2mol/RDKit's 91.53% by a margin of 4 percentage points. The ROC-AUC metric further highlights MolBar's superior discriminative ability, with a score of 95.94% compared to 90.74% for xyz2mol/RDKit. MolBar has a clear advantage in precision, achieving 97.60% versus 88.85% for xyz2mol/RDKit. This difference of nearly 9 percentage points indicates that MolBar is more effective at correctly identifying non-rotatable bonds, minimizing false positives. The recall metric shows that xyz2mol/RDKit achieves slightly higher sensitivity with a score of 97.11% compared to MolBar's 94.88%. However, this higher recall comes at the cost of substantially reduced precision, indicating that xyz2mol/RDKit captures more true non-rotatable bonds but also generates more false positives. The F1-score, which balances precision and recall, further underscores MolBar's superiority with a score of 96.22% compared to 92.80% for xyz2mol/RDKit. The specificity metric

Table 5.3: Comparison of classification performance between MolBar's MPNN classifier and xyz2mol/RDKit for distinguishing between rotatable and non-rotatable bonds in a challenging subset of 100 transition metal complexes from the hold-out test set. The ground truth labels are derived from GFN2-xTB WBO calculations with a threshold of 1.25. MolBar exhibits superior performance across all metrics except recall, indicating a more balanced classification ability.

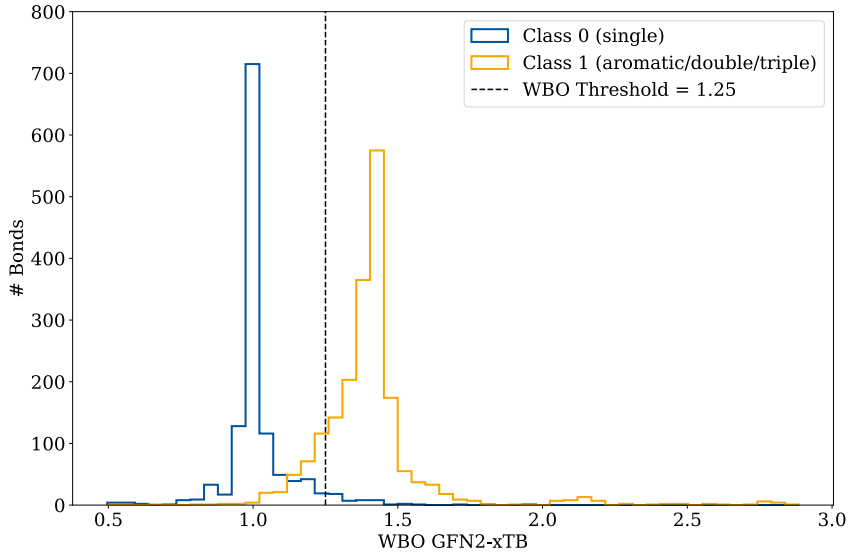
Metric	MolBar	xyz2mol/RDKit
Accuracy	95.81%	91.53%
ROC-AUC	95.94%	90.74%
Precision	97.60%	88.85%
Recall	94.88%	97.11%
F1-score	96.22%	92.80%
Specificity	97.00%	84.37%

further underlines MolBar's advantage: at 97.00% compared to 84.37% for xyz2mol/RDKit, it demonstrates a superior ability to correctly identify rotatable bonds (true negatives). It is important to note that this comparison between MolBar and xyz2mol/RDKit is not fully equitable. MolBar, unlike xyz2mol, was specifically trained with this threshold to distinguish between rotatable and non-rotatable bonds. xyz2mol/RDKit's goal is to generate Lewis structures (or more specifically SMILES codes) from 3D coordinates and not assign rotatability. This highlights the distinction between general bond classification and the specific task of predicting rotatability, emphasizing the need for a model tailored to the specific goal. A distinction between single versus aromatic/double/triple bonds as with xyz2mol/RDKit does not reflect the threshold necessary to predict a rotational barrier of 20 kcal/mol.

Chemically more interesting and less abstract than an analysis over an entire dataset is the analysis of single molecules and the number of misclassified bonds per molecule. Figure 5.7 shows the distribution of false positive (right) and false negative (left) bonds per transition metal complex. While the classification errors depicted in Figure 5.7 may initially seem substantial, their impact is attenuated upon examining the corresponding WBO values for misclassified instances. MolBar outperforms xyz2mol/RDKit in false-positive classification, demonstrating a pronounced peak at zero false positives and a mean of only 0.42 false positives per molecule ($\sigma = 0.91$). In contrast, xyz2mol/RDKit exhibits a broader error distribution, yielding a mean of 2.21 false positives per molecule ($\sigma = 3.05$). For false negatives, MolBar shows a mean of 0.93 errors per molecule ($\sigma = 1.83$), compared to xyz2mol/RDKit's lower mean of 0.53 ($\sigma = 1.17$). This supports the findings in Table 5.3 that xyz2mol/RDKit has higher recall but lower precision compared to MolBar. With both distributions, MolBar overall demonstrates superior performance, with fewer misclassifications per molecule and a more concentrated error distribution. Analyzing the percentiles yields that for MolBar, 66% of the 135 misclassifications occur in a WBO range between 1.21 and 1.31, which is close to the 1.25 threshold, indicating that these errors are likely due to borderline cases where the model's confidence is lower. For xyz2mol/RDKit, the misclassifications are more widely distributed across a broader range of WBO values. Here, 66% of the 274 misclassifications range from 1.10 to 1.27, suggesting that



(a) MolBar



(b) xyz2mol/RDKit

Figure 5.6: Distribution of reference WBO, stratified by predicted bond classes for the MPNN classifier from MolBar (left) and xyz2mol/RDKit (right) on a challenging subset of 100 transition metal complexes. The x-axis represents the ground-truth values WBO obtained from GFN2-xTB calculations, while the y-axis indicates the number of bonds. The vertical dashed line marks the threshold of 1.25 that separates rotatable from non-rotatable bonds. Compared to the results shown in Figure 5.5, the separation achieved with MolBar shows comparable performance, albeit with a slight decline due to the inherent complexity of the subset under consideration. xyz2mol/RDKit tends to overestimate the WBO, resulting in some bonds being classified as non-rotatable (orange) even though they have a WBO below 1.25.

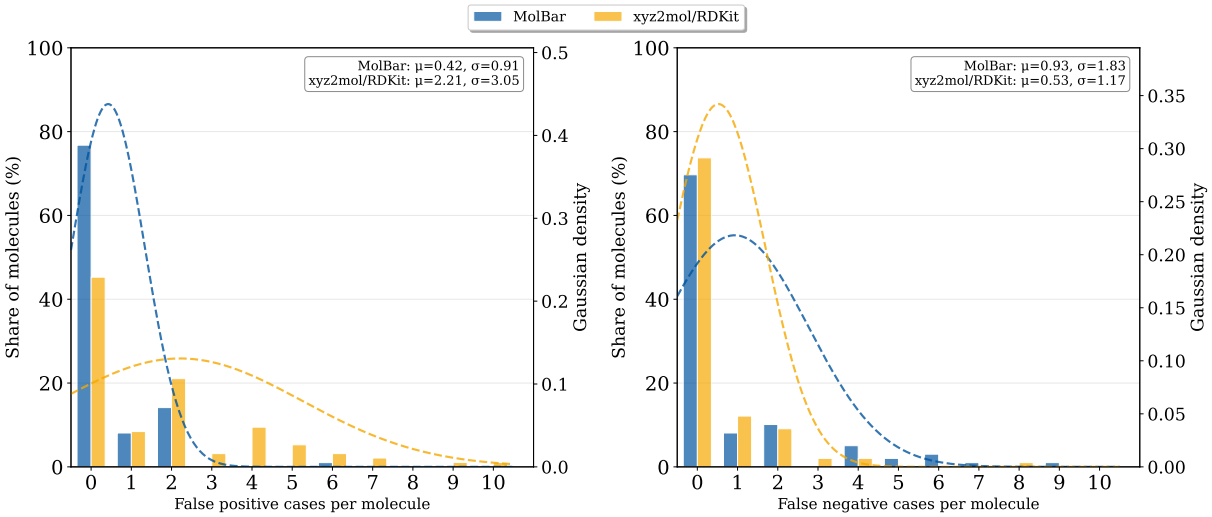
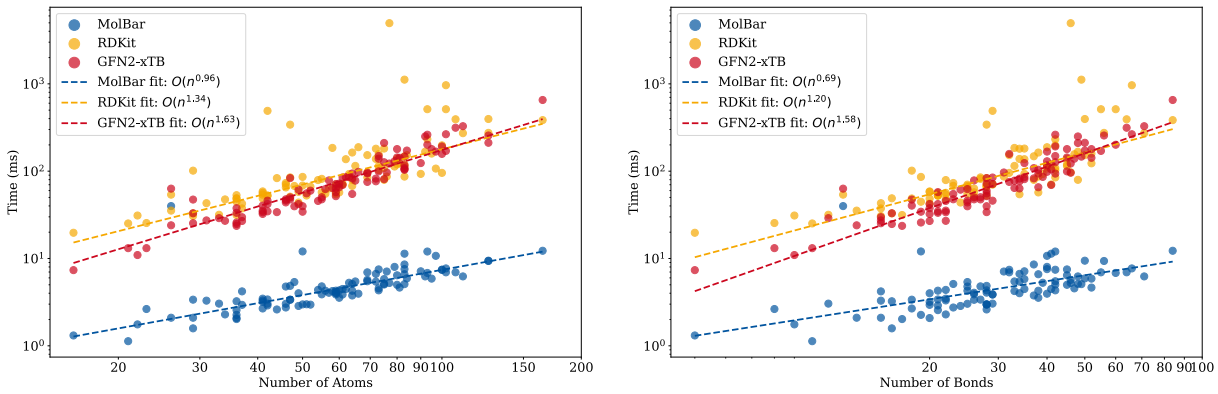


Figure 5.7: Distribution of classification errors per molecule for MolBar's MPNN classifier (blue) and xyz2mol/RDKit (orange) on the subset of 100 transition metal complexes. The left panel shows false positive cases (bonds incorrectly predicted as non-rotatable), while the right panel displays false negative cases (bonds incorrectly predicted as rotatable). The x-axis represents the number of misclassified bonds per molecule, and the y-axis indicates the number of molecules with that count of misclassifications. MolBar demonstrates superior performance in false-positive classification compared to xyz2mol/RDKit. For false-negative classification, MolBar shows a slightly higher mean number of errors per molecule compared to xyz2mol/RDKit. Overall, MolBar's misclassifications are more tightly distributed around the threshold of 1.25 (66% percentiles between 1.21 and 1.31), while xyz2mol/RDKit's misclassifications are more widely distributed across a broader range of WBO values (66% percentiles between 1.10 and 1.27).

the errors are not confined to borderline cases but may also occur in instances where the distinction between rotatable and non-rotatable bonds should be clearer. Thus, MolBar emerges as a more robust classifier, particularly in scenarios where discriminative clarity is limited.

Beyond predictive performance, computational efficiency remains a critical practical consideration for large-scale applications. Figure 5.8 presents a logarithmic scaling analysis of the execution time for WBO prediction, plotted as a function of molecule size (the number of atoms and bonds) for the MPNN classifier from MolBar, xyz2mol/RDKit, and the reference method GFN2-xTB. The left diagram illustrates computational cost as a function of atom count, while the right diagram depicts scaling with respect to bond count. Both diagrams reveal power-law relationships fitted to the empirical data points, with the exponents reflecting the underlying algorithmic complexity. The MolBar classifier demonstrates nearly linear scaling behavior, with fitted exponents of approximately 0.96 for atoms and 0.69 for bonds. This advantageous scaling arises from the message-passing architecture, where the computational cost grows linearly with graph size due to sparse matrix operations. In contrast, xyz2mol/RDKit exhibits a steeper scaling, with adjusted exponents of 1.34 for atoms and 1.20 for bonds, indicating superlinear complexity. More significantly, the prefactors for xyz2mol/RDKit (3.69×10^{-4} for atoms, 1.51×10^{-3} for bonds) are approximately 4 and 3.5 times larger, respectively, compared to MolBar (8.94×10^{-5} for atoms, 4.26×10^{-4} for bonds). The computational cost of xyz2mol/RDKit aligns closely with that of the reference calculation GFN2-xTB (9.77×10^{-5} for

atoms, 3.32×10^{-4} for bonds), which, despite having larger scaling exponents, features smaller prefactors. For molecules comprising 50 to 100 atoms, xyz2mol/RDKit requires execution times on the same order of magnitude as GFN2-xTB, effectively eliminating any computational advantage over direct WBO calculation. The graph neural network approach of MolBar amortizes computational costs during training, achieving inference times approximately an order of magnitude faster than GFN2-xTB across the entire range of molecule sizes studied.



(a) Scaling with number of atoms.

(b) Scaling with number of bonds.

Figure 5.8: Log-log analysis of the scaling of computational costs for the MPNN classifier from MolBar (blue), xyz2mol/RDKit (orange), and reference calculations with GFN2-xTB (red) as functions of molecule size. The scatter points represent individual molecules from the test subset, while the dashed lines indicate fitted power-law relationships $t = a \cdot n^b$, where n denotes the molecule size and b the scaling exponent. MolBar exhibits sublinear to linear scaling with exponents of $b = 0.96$ (atoms) and $b = 0.69$ (bonds), with prefactors of $a = 8.94 \times 10^{-5}$ (atoms) and $a = 4.26 \times 10^{-4}$ (bonds), consistent with the efficient message-passing architecture of graph neural networks using sparse matrix operations. In contrast, xyz2mol/RDKit shows superlinear scaling with $b = 1.34$ (atoms) and $b = 1.20$ (bonds), with substantially larger prefactors of $a = 3.69 \times 10^{-4}$ (atoms) and $a = 1.51 \times 10^{-3}$ (bonds), while GFN2-xTB scales as $b = 1.63$ (atoms) and $b = 1.58$ (bonds) with prefactors of $a = 9.77 \times 10^{-5}$ (atoms) and $a = 3.32 \times 10^{-4}$ (bonds). The substantially larger prefactors of xyz2mol/RDKit result in execution times that approach those of GFN2-xTB for molecules exceeding 50 atoms, undermining its utility as a fast alternative. MolBar maintains approximately an order of magnitude speed advantage over GFN2-xTB across all evaluated molecule sizes.

The combined results from the classification performance (Table 5.3), the WBO distribution alignment (Figure 5.6), and the computational efficiency (Figure 5.8) highlight the necessity of a specialized, task-specific model for predicting rotatability. While xyz2mol/RDKit excels in assigning chemically meaningful bond orders to SMILES codes, it proves unsuitable for classifying bonds as rotatable or non-rotatable. This limitation becomes evident in the significant overlap of the WBO distribution in Figure 5.6, where the classifications from xyz2mol/RDKit fail to perfectly correlate with the 1.25 threshold distinguishing rotatable from non-rotatable bonds. Additionally, the comparable computational cost between xyz2mol/RDKit and GFN2-xTB renders the former impractical as a replacement for the latter in production workflows. With equivalent computational effort, direct WBO calculation via GFN2-xTB delivers fundamental accuracy without introducing classification artifacts. The architecture of MolBar addresses both challenges simultaneously: (i) it is explicitly trained to predict rotatability rather than bond type, yielding superior precision and specificity for the downstream task, and (ii) it

achieves this with near-linear scaling, enabling practical deployment in high-throughput scenarios. Thus, the developed method proves sufficient for integration within the rigidity matrix in Equation 4.7.

5.4 Rigidity Matrix: Steric Hindrance Model

5.4.1 Introduction

Given the sets of electronically hindered bonds $M_{\text{electronic}}$ (Section 5.3) and bonds in rings C^s (Section 2.3), the final component of the rigidity matrix $\mathcal{I}$ (Equation 4.7) is the set of sterically hindered bonds M_{steric}.[154] Sterically hindered bonds are single bonds with high rotational barriers due to steric interactions between substituents. These barriers enable the formation of atropisomers, which can be isolated and characterized as distinct stereoisomers. While ortho-substituted biaryl systems, such as biphenyl derivatives, are classic examples,[197] atropisomerism also occurs in acyclic systems like N-aryl amides and various heterocycles.[198] The study of atropisomerism originated in the early 20$^{\text{th}}$ century. In 1922, Christie and Kenner provided the first empirical evidence by isolating two stable conformers of 6,6′-dinitro-2,2′-diphenic acid via fractional crystallization.[199] They observed two distinct forms with different physical properties, confirming their stereochemical non-equivalence. Their reasoning suggested that both aromatic rings shared a rotational axis but were not coplanar, rendering them non-superimposable.

In 1933, Richard Kuhn introduced the term "atropisomer", derived from the Greek "atropos" ("without turn"), to describe stereoisomers resulting from impeded rotation. However, the principles of axial chirality, the primary form of chirality in atropisomerism, were established earlier. Van 't Hoff first theorized axial chirality in 1875, specifically for allenes. Maitland and Mills later validated this concept in 1935 by synthesizing and characterizing optically active allene derivatives. While axial chirality in allenes arises from their unique geometry, biaryl enantiomers formed a separate class. Researchers noted that bulky ortho substituents could sterically hinder rotation about the aryl-aryl bond.

In 1958, Mislow provided a rigorous definition of atropisomerism as "stereoisomerism produced by inhibition of free rotation about a single bond". This is synonymous with "rotational isomerism".[200] Ōki later advanced this understanding by introducing a kinetic framework.[74] He proposed that molecules with a half-life for rotational conversion exceeding 1,000 seconds under ambient conditions be classified as atropisomers, corresponding to an energy barrier of approximately 22 kcal/mol. LaPlante et al. further refined this concept for pharmaceutical development by categorizing atropisomers into three classes based on their racemization half-life at 37 °C.[113] Class 1 atropisomers, with a half-life under 60 seconds ($t_{1/2} < 60$ s), exhibit rapid conversion and effectively achiral behavior. Class 2 atropisomers, with half-lives between 60 seconds and 4.5 years (60 s $< t_{1/2} < 4.5$ years), are isolable and observable by nuclear magnetic resonance (NMR) but may lack long-term stability. Class 3 atropisomers, with half-lives exceeding 4.5 years ($t_{1/2} > 4.5$ years), are considered completely stable and suitable for development as single enantiomer drugs. Class 2 atropisomers present a particular challenge in drug development due to their intermediate stability. The risk of racemization

during manufacturing or storage can compromise stereochemical identity, leading to regulatory concerns. Medicinal chemists are therefore advised to modify such compounds to achieve either "stable" (class 3) or "non-chiral" (class 1) status.[201] These classifications underscore the broad impact of atropisomerism and induced chirality on drug development. The significance of enantiomers for drug safety and efficacy has been recognized for decades, with thalidomide serving as a symbolic example of the risks associated with neglecting stereoisomers.[202] Sotorasib, a covalent $KRAS^{G12C}$ inhibitor approved in 2021, exemplifies atropisomerism.[203] Its ortho-substituted biaryl motif forms a configurationally stable pair of atropisomers, designated as M and P forms. These forms correspond to left- and right-handed helical arrangements of the biaryl system. The M form exhibits a tenfold higher affinity for the KRAS protein pocket, enhancing the inhibitor's potency. Lesinurad, approved in 2015 for treating high blood uric acid levels associated with gout, provides another example.[204] Post-approval studies revealed that the (+) atropisomer is approximately four times more effective at URAT1. Surveys indicate that about 15% of all FDA-approved small molecule therapeutics—and nearly 30% of recent approvals—contain at least one rapidly interconverting atropisomer axis. Even class 1 axes, which racemize rapidly, can significantly influence biological activity.

While atropisomerism's importance in medicinal chemistry is increasing, its integration into standard cheminformatics tools remains limited. BIOVIA CT file formats encode atropisomeric stereochemistry using wedged and hashed bonds.[205] CDD Vault employs a similar annotation strategy.[206] `RDKit`, starting with version 2024.03.1, now supports explicit atropisomeric bond designation using M or P notation.[86] Here, an atropisomeric bond is defined as one where each atom has at least one adjacent single or aromatic bond with a specified direction. `RDKit` does not attempt to actually classify bonds as rotationally hindered or not, but rather classifies any bond that meets the aforementioned criteria as atropisomeric. Molecular line notations face greater challenges. SMILES, for instance, lacks formal syntax for axial chirality, though extensions allow atropisomeric annotations.[207,208] InChI can represent systems like allenes and cumulenes but lacks a universal method for atropisomeric configurations.[39] The InChI Trust is addressing this in future releases.[87] Group SELFIES attempts to capture axial stereochemistry via group-level tokens, which could accommodate atropisomerism.[92] A persistent limitation across these approaches is the requirement for manual specification of atropisomeric bonds, making them impractical for large-scale datasets. Identification typically depends on experimental data or computationally intensive barrier evaluations.

Rapid identification of atropisomeric bonds, ideally within milliseconds, remains a key challenge in cheminformatics. Theoretically, determining atropisomerism requires evaluating the MERP for bond rotation, with the transition state defining the rotational barrier. In biaryl systems, steric repulsion from ortho-substituents often dominates, simplifying the problem. However, complex systems like tertiary aromatic amides may exhibit coupled torsional motions that lower the effective barrier.[209] Intermolecular interactions, such as hydrogen bonding, can also stabilize conformations and increase barriers. Given MolBar's operational timescale (tenths to hundreds of milliseconds), any detection method must stay within the computational

budget of identifier calculation. While DFT scans provide accurate barriers,[210] their computational cost precludes real-time use. Even semiempirical methods like GFN2-xTB[46] are too slow for on-the-fly processing. Studies have quantified the impact of ortho-substituent steric radii on rotational barriers in biaryl systems. Bott et al. showed that barrier height correlates with the effective van der Waals (vdW) radii of substituents,[211] offering a quantitative basis for atropisomer stability trends. Electron-withdrawing or -donating groups further influence conjugation and barrier energy.[212]

5.4.2 Steric Hindrance Model

This section introduces a physical model that classifies bonds as either sterically hindered or unhindered. This model is based on the observation that the vdW radii of substituents directly affect rotational barriers. It evaluates the probability of steric hindrance, denoted as p_{ij}^{steric}, for a bond (i, j). For calculating this probability, the model considers the 3D conformation of the two fragments connected by the candidate bond using a distance-dependent potential to estimate steric repulsion. The model is designed to be computationally efficient, enabling rapid classification within the MolBar framework.

Atropisomerism most commonly arises at single bonds between sp^2-hybridized atoms, especially in biaryl systems where ortho substituents restrict rotation. Thus, the classifier targets sp^2-sp^2 single bonds that are not embedded in ring systems. Bonds within ring systems (C^S) or those with electronic hindrance ($M_{\text{electronic}}$) are excluded, as their rotation is already constrained by other mechanisms. Rare cases, such as chiral sp^3-sp^3 axes in triptycenes[213] and chiral sp^2-sp^3 axes in 9-arylfluorenes,[214] are omitted to maintain focus on chemically prevalent motifs. Atropisomeric bonds in macrocycles are also excluded. The prevalence of atropisomerism in biaryl systems suggests a connection to the idea of MolBar using rigid fragments. Each aryl ring represents a distinct fragment, and the potential atropisomeric bond links these two fragments. Ignoring atropisomerism in macrocycles, potential atropisomeric axes are single bonds that are not part of ring systems by definition. Hence, any atropisomeric bond connects two fragments in the MolBar model.

After fragmenting the molecular structure using $M_{\text{electronic}}$ and C^S, the classifier assesses each sp^2-sp^2 bond connecting two fragments to determine if it is sterically hindered. This approach reduces the analysis to the 3D conformation of two neighboring fragments. The simplification limits the scope to local steric interactions, neglecting distant substituents or flexible side chains. The underlying assumption is that, under equilibrium conditions, these side chains will reorder in a way that offers a minimum energy path for the bond rotation, which is implicitly assumed within the locality approximation. While this locality approximation assumes that the rotational barrier depends primarily on the bond's immediate environment, it provides a computationally efficient framework. A comprehensive treatment of the entire molecule would demand extensive conformational sampling and energy calculations, which

are prohibitive for high-throughput applications in MolBar. The two-fragment system offers a pragmatic simplification for atropisomerism. It emphasizes the essential steric interactions that govern rotational barriers, with clearly defined limitations.

Incorporating this classifier into the rigidity matrix $\mathcal{I}$ (Equation 4.7) enables MolBar to describe atropisomerism directly from 3D coordinates without manual intervention or extensive computational resources. The model is deliberately lightweight and more physically motivated compared to the electronic hindrance classifier in Section 5.3 using a neural network due to limited data availability. Using a distance-dependent potential, the steric repulsion for a given conformer m of the two-fragment system is calculated by summing over all pairs of atoms within those fragments:

$$E_{\text{rep}}^{(m)} = \sum_{k \neq l} \left(\frac{r_{kl}^{\text{vdW}}}{r_{kl}^{(m)}} \right)^n , \tag{5.30}$$

where $r_{kl}^{(m)}$ is the distance between atoms k and l in conformer m, and r_{kl}^{vdW} is the sum of covalent radii for atoms k and l. n is a trainable parameter that determines the steepness of the potential, with larger values leading to sharper repulsion at short distances. The covalent radii are treated as trainable parameters, enabling the model to adapt them to the specific classification task. The vdW distance between atoms k and l is then defined as the sum of their effective covalent radii:

$$r_{kl}^{\text{vdW}} = r_k^{\text{cov}} + r_l^{\text{cov}}. \tag{5.31}$$

By multiplying a rotation matrix with the 3D fragment coordinates, a set of conformers m is generated through scanning the candidate bond without relaxation. This yields a potential energy profile E_{rep}, which is used to identify the conformer with the lowest steric energy and the second-highest local maximum, corresponding to the minimum and relevant transition state, respectively. The conformer with the lowest steric energy is identified as:

$$m_{\text{min}} = \underset{m}{\text{argmin}}\, E_{\text{rep}}^{(m)}. \tag{5.32}$$

The second-highest local maximum is used as a transition state guess. It is defined as:

$$\begin{aligned} m_{\text{2nd max}} = & \underset{m \in \mathcal{M}_{\text{loc}} \setminus \{m_{\max}\}}{\text{argmax}} E_{\text{rep}}^{(m)}, \\ & \text{where} \quad \mathcal{M}_{\text{loc}} = \{m : E_{\text{rep}}^{(m-1)} < E_{\text{rep}}^{(m)} > E_{\text{rep}}^{(m+1)}\}, \\ & \text{and} \quad m_{\max} = \underset{m \in \mathcal{M}_{\text{loc}}}{\text{argmax}}\, E_{\text{rep}}^{(m)}. \end{aligned} \tag{5.33}$$

Here, $\mathcal{M}_{\text{loc}}$ is the set of local maxima in the steric energy profile, and $m_{\max}$ is the index of the global maximum. By excluding $m_{\max}$, the second-highest local maximum is selected, which is more likely to represent a realistic transition state for rotation. The steric barrier height is then:

$$\Delta E_{\text{rep}} = E_{\text{rep}}^{(m_{\text{2nd max}})} - E_{\text{rep}}^{(m_{\text{min}})}. \tag{5.34}$$

To allow for a more flexible model, the steric barrier is used inside a simple logistic regression model with two trainable parameters, a_0 and a_1. That is, the computed steric barrier ΔE_{rep} is linearly transformed to a value y_{ij}^{steric}:

$$y_{ij}^{\text{steric}} = a_1 \cdot \Delta E_{\text{rep}} + a_0. \tag{5.35}$$

The model applies a sigmoid function to the output to yield a probability value p_{ij}^{steric} that indicates whether the bond is sterically hindered:

$$p_{ij}^{\text{steric}} = \frac{1}{1 + e^{-y_{ij}^{\text{steric}}}}. \tag{5.36}$$

Using a defined threshold ω_{steric}, the model classifies a bond (i, j) as sterically hindered or unhindered. A bond is included in the set of sterically hindered bonds M_{steric} if its predicted probability exceeds the threshold:

$$M_{\text{steric}} = \{(i, j) \mid \forall (i, j) \in E(G) \wedge p_{ij}^{\text{steric}} > \omega_{\text{steric}}\}. \tag{5.37}$$

5.4.3 Computational Protocol and Details

The development of a steric hindrance classifier requires reliable reference data for both training and validation. Experimental rotational barriers are frequently reported in the literature, often determined through dynamic nuclear magnetic resonance (DNMR)[198,211,215–217] or chiral high-performance liquid chromatography (HPLC).[218,219] However, these measurements tend to focus on stable atropisomers, which biases datasets toward higher barriers. To address this limitation, a computational protocol is essential for estimating rotational barriers across a broader range of bonds. This approach ensures both consistency and comprehensiveness in the reference data. The following sections describe and benchmark a computational workflow against experimental results. The validated protocol is subsequently applied to generate reference data for training and testing the classifier. Figure 5.9 provides an overview of the computational workflow.

To generate meaningful energy profiles, the two rigid fragments used for prediction must be chemically saturated. Hydrogen atoms or the respective terminal element are added to maintain the local electronic structure of the full molecule's substructure (Figure 5.9, step 2b). The resulting substructures are validated using `RDKit`[86] to ensure chemical plausibility and exclude unrealistic bonding scenarios. A relaxed dihedral scan is then performed on each validated fragment pair using `xtb` (version 6.7.1) with the GFN2-xTB method[46] (Figure 5.9, step 3). The torsional angle is incremented in 1° steps over a full 360° rotation. Each generated geometry is evaluated using the MolBar bond assessment routine (Section 5.2) to detect bond-breaking events. Scans exhibiting chemical instability are discarded. However, clashing atoms—particularly those involving bulky substituents—are retained, as such interactions are

relevant to atropisomerism. At each step, all other internal degrees of freedom are relaxed through constrained optimization with the GFN2-xTB method.[46]

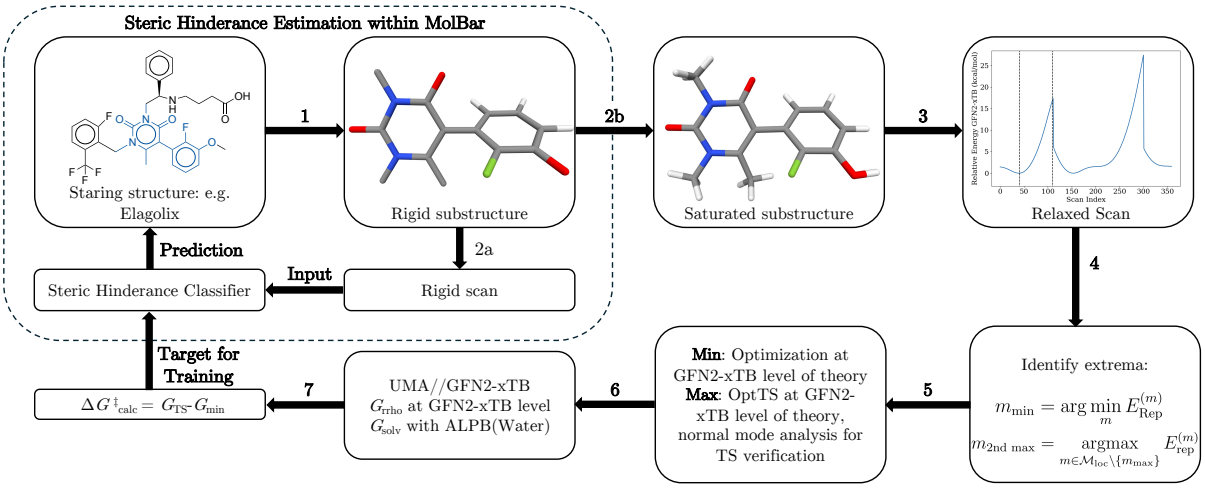


Figure 5.9: Computational protocol for calculating targets and features for the steric hindrance classifier. It determines rigid scan conformers (features) and computes the rotational barrier (target) using electronic structure methods. First, candidate bonds between rigid MolBar fragments are selected (1). Then a rigid scan using the plain unsaturated fragments is performed around the candidate bond (2a) and the two fragments are saturated (2b). A relaxed dihedral scan at the GFN2-xTB level of theory is performed next (3), identifying minima and maxima (4). The minimum and transition state structures are optimized (5), single-point energies using the UMA MLIP and thermodynamic corrections are computed (6) and the rotational barrier is calculated as the free energy difference between the transition state and the minimum (7). The unsaturated rigid scan conformers serve as input for the classifier model while the calculated rotational barriers provide reference data. The steps enclosed in the dashed box are carried out as part of the hindrance estimator inside MolBar, while the steps outside the box are performed as part of the reference data generation for training and validation of the classifier. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

Minima and maxima were identified from the relaxed dihedral angle scan (Figure 5.9, step 4). The global minimum corresponds to the most stable conformer on the torsional energy surface. For the initial transition state estimate, the second-highest maximum was selected instead of the global maximum, as this path presents a more probable, lower-energy barrier. The minimum was optimized without constraints using GFN2-xTB (Figure 5.9, step 5). The transition state candidate was optimized with the `ORCA` quantum chemistry suite (version 6.0.1).[40] To precisely locate first-order saddle points along the torsional coordinate, the `OptTS` keyword was used with `TightOpt` and `TightSCF` for stricter convergence. The `SlowConv` option enhanced SCF stability near saddle points. The desired torsional degree of freedom was specified using the `%geom TS_Mode` block. Transition state validity was confirmed by verifying a single imaginary frequency corresponding to the torsional mode, calculated with the `NumFreq` keyword. The transition state mode was analyzed for significant changes in the relevant dihedral angle, with validation confirmed if the change exceeded a predefined threshold.

To refine the rotational barriers, single-point electronic energies for both the minimum and

transition state structures were computed using the MLIP variant UMA-S[65] (Figure 5.9, step 6). UMA utilizes an equivariant graph neural network (EGNN)[182] with a mixture of linear experts (MoLE) architecture. The model was trained on reference data at the ωB97M-V/def2-TZVPD[220,221]//GFN2-xTB level. UMA was chosen for its computational efficiency and accuracy in predicting conformational energies, achieving a mean absolute error of 5.76 meV (0.133 kcal/mol) in the ΔE metric on the OMol25 benchmark.[58] Geometry optimizations were performed with GFN2-xTB[46] rather than UMA-S due to its superior convergence for transition states in `ORCA`[40] (version 6.0.1). Thermodynamic corrections were derived from GFN2-xTB frequency analysis under the rigid-rotor harmonic-oscillator (RRHO) approximation.[46] Solvation effects were included via the analytical linearized Poisson-Boltzmann (ALPB) implicit solvation model,[222] calculating free energy differences between gas phase and solvated structures. Frequency calculations were conducted at 298.15 K unless specified otherwise (e.g., Table 5.4). Free solvation energies, δG_{solv}, were computed for water based on gas-phase geometries. The total free energy of each structure was defined as:

$$G = E_{\text{elec, UMA}} + G_{\text{RRHO, GFN2-xTB}} + \delta G_{\text{solv, ALPB}} \tag{5.38}$$

The rotational barrier was calculated as the difference in free energy between the transition state and the global minimum, as illustrated in Figure 5.9 (step 7):

$$\Delta G^{\ddagger}_{\text{calc.}} = G_{\text{TS}} - G_{\text{min}} \tag{5.39}$$

This calculated free energy difference, $\Delta G^{\ddagger}_{\text{calc.}}$, was used as the reference value for both training and validation of the steric hindrance classifier. The input features for the classifier were derived from the unsaturated rigid scan conformers.

5.4.4 Validation of the Computational Protocol using Experimental Data

To evaluate the accuracy of the computational protocol and the validity of the locality approximation, a benchmark dataset was compiled from the literature. This set consists of atropisomeric compounds with known experimental rotational free energy barriers, detailed in Table 5.4. The barriers of these selected molecules range from 10.6 to 28.9 kcal/mol, predominantly lying between 15 and 25 kcal/mol. The dataset includes instances where the MolBar fragments constitute the full molecule, alongside more complex scaffolds that feature additional substituents or flexible side chains. The specific substructures employed for the barrier calculations are marked in blue in Table 5.4. Figure 5.10 presents a parity plot correlating the experimental and computed rotational free energy barriers ($\Delta G^{\ddagger}$) for this benchmark set. The established MolBar threshold of 20 kcal/mol, which distinguishes conformational flexibility from stable molecular identity, is thoroughly covered, with multiple compounds situated near this critical limit. This distribution enables a rigorous assessment of the protocol's classification capability for predicting whether bonds are rotationally hindered. In the plot, this threshold

is marked by a horizontal line, and the surrounding grey band denotes a ±2 kcal/mol margin of error, reflecting the inherent uncertainty zone where classification is difficult due to typical electronic structure calculation deviations.

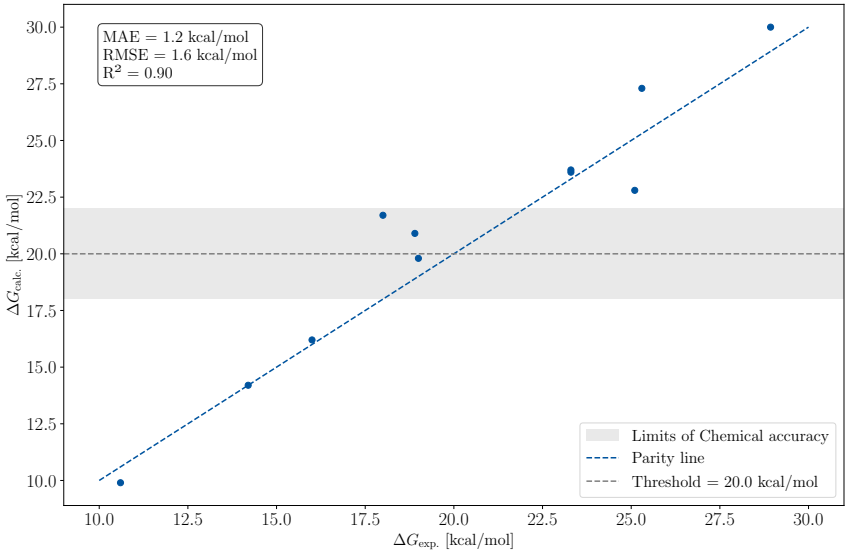
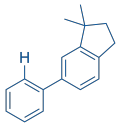
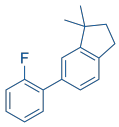
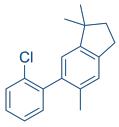
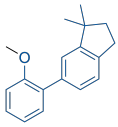
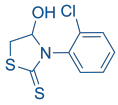
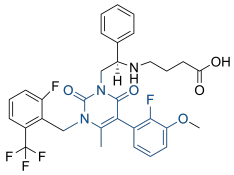
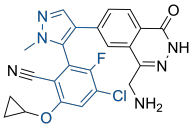
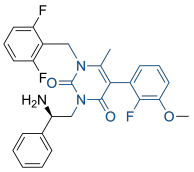
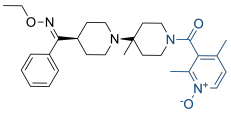
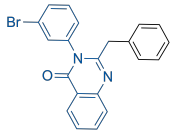
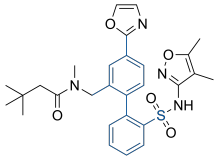


Figure 5.10: Parity plot between experimental and computed rotational barriers computed at UMA-S[58]//GFN2-xTB[46] level of theory for a benchmark set of atropisomeric compounds. Thermodynamic corrections were calculated using RRHO approximation at the GFN2-xTB level[46] and solvation effects were included via the ALPB[222] implicit solvent model. The horizontal line at 20 kcal/mol represents the threshold used in MolBar to classify bonds as rotationally hindered or not. The grey region highlights barriers within ±2 kcal/mol of this threshold, indicating where any classification is highly uncertain due to common electronic structure theory errors. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

The protocol demonstrates strong quantitative agreement with experimental data, achieving a mean absolute error of 1.2 kcal/mol and a root mean square error of 1.6 kcal/mol. Despite the use of isolated, saturated substructures, the method yields a coefficient of determination (R^2) of 0.90, consistent with expectations for electronic structure theory accuracy. Most compounds are correctly classified relative to the 20 kcal/mol threshold. However, discrepancies occur near this boundary, where minor errors can alter classification. For example, compound **3** (18.0 kcal/mol) and BrPh-QZ-Bn (**10**, 18.9 kcal/mol) are slightly overestimated at 21.7 and 20.9 kcal/mol, respectively, leading to false positives.

Molecules with rotational barriers between 18 and 22 kcal/mol should be treated as borderline cases, requiring cautious interpretation. Outside this range, the protocol performs robustly, supporting its application in both quantitative regression and binary classification for atropisomer screening. The benchmark analysis confirms the efficacy of the computational protocol (cf. Figure 5.9, steps 2–7) and validates the locality approximation, providing confidence in generating reliable reference values for training a lightweight atropisomer model that can be applied on-the-fly within MolBar without UMA or GFN2-xTB calculations.

Table 5.4: Benchmark of experimental and computed rotational barriers for a set of atropisomeric compounds, using isolated MolBar fragments (colored in blue), at the UMA-S//GFN2-xTB level[46,65] with RRHO thermodynamic corrections[46] and ALPB[222] solvation.

Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{exp.}}$	$\Delta G^{\ddagger}_{\text{calc.}}$	Solv.	T [°C]	Technique	Source
	**1**	10.6	9.9	DMSO	67	DNMR	[211]
	**2**	14.2	14.2	DMSO	67	DNMR	[211]
	**3**	18.0	21.7	DMSO	67	DNMR	[211]
	**4**	16.0	16.2	DMSO	67	DNMR	[211]
	**5**	19.0	19.8	DMSO	25–124	DNMR	[215]
	**6** (Elagolix)	23.3	23.7	$CDCl_3$	37	DNMR	[198]
	**7** (MRTX1719)	28.9	30.0	TCE	25	^{19}F NMR	[223]
	**8** (NBI-42902)	23.3	23.6	DMSO	25	DNMR	[216]
	**9** (SCH351125)	25.3* ($t_{1/2}$ = 5.5 h)	27.3	water	37	HPLC, $t_{1/2}$	[218]
	**10** (BrPh-QZ-Bn)	18.9	20.9	nitrobenzene	94	DNMR	[217]
	**11** (BMS-207940)	25.1* ($t_{1/2}$ = 15.2 h)	22.7	water	37	HPLC, $t_{1/2}$	[219]

*Calculated using $\Delta G^{\ddagger} = RT \ln\left(\frac{k_B T t_{1/2}}{h \ln 2}\right)$

5.4.5 Dataset

Developing a robust steric hindrance model demands a dataset that adequately represents both atropisomeric and non-atropisomeric molecules. Given the scarcity of suitable benchmarks, the beta-secretase 1 (BACE) collection (1,513 drug-like inhibitors from MoleculeNet[224,225]) was selected as a foundation. This dataset was systematically expanded to include literature-known atropisomeric biaryl and binaphthyl derivatives, ensuring a balanced basis for training and evaluation. A full inventory of these supplementary structures is detailed in the Appendix (Section A.2). The data is categorized into successful, converged barrier calculations (Tables B.1 to B.9) and those that failed to converge (Tables B.10 to B.15). While non-converged instances were omitted from the training set, they are documented here for complete transparency. The computational protocol (Figure 5.9) was applied to this combined dataset to generate reference rotational barriers for all candidate bonds between rigid MolBar fragments. This process produced 1,143 connected fragments with converged barrier calculations. Bonds with a WBO exceeding 1.2 were excluded, as electronic effects dominate in these cases. To reduce redundancy, filtering by MolBar identifier yielded 561 unique connected fragments. Of these, 33 fragments (5.9%) were classified as atropisomeric based on their computed rotational barriers.

5.4.6 Model Training and Evaluation

The steric hindrance classification model was optimized using `PyTorch`[190] (version 2.9.1) and `SciPy`[135] (version 1.17.0). The Adam optimizer was initialized with a learning rate of 10^{-2}. In addition to the three core parameters (n, a_1, and a_0), the van der Waals radii of carbon and bromine were included as trainable parameters, as exploratory analyses indicated their significance. Initial values were set to $n = 3$, $a_1 = 3$, and $a_0 = -6$, with atomic radii initialized using tabulated values from Pyykkö and Atsumi.[153] A dynamic learning rate scheduler reduced the step size when the validation loss plateaued. To address class imbalance between low (< 20 kcal/mol) and high ($\geqq 20$ kcal/mol) rotational barriers, a weighted binary cross-entropy loss was applied. Early stopping was implemented to halt training if validation loss did not improve over five consecutive epochs. The model parameters at the point of minimum validation loss were retained for final evaluation. A three-fold CV framework was employed to train and evaluate the model using 549 unique connected molecular fragments. This approach balanced training data sufficiency with reliable performance estimation, especially given the dataset's class imbalance. As with the electronic hindrance classifier, multiple data-splitting strategies were tested to prevent data leakage and ensure realistic generalizability. These include random, scaffold-based, Butina-based, and UMAP-based splits.

Random Split

The random split strategy assigns molecules to training and test sets without regard to structural relationships. Although simple to implement, this method risks including structurally similar compounds in both sets, which may inflate performance estimates.

Scaffold Split

The scaffold-based split groups molecules by their Bemis–Murcko scaffolds using `RDKit`.[86] All molecules sharing a scaffold are assigned to either the training or test set.[193] This approach ensures separation based on core molecular frameworks, so that the model is evaluated on structurally distinct compounds.

Butina Split

The Butina-based split generates Morgan fingerprints via `RDKit`[86] and applies the Butina clustering algorithm to group structurally similar molecules.[226] Molecules within the same cluster are assigned to the same split, increasing structural dissimilarity between training and test sets by leveraging fingerprint similarity rather than scaffold identity alone.

UMAP-based Split

The UMAP-based strategy embeds Morgan fingerprints into a two-dimensional space using the UMAP algorithm with parameters `n_neighbors = 7` and `min_dist = 0.1`, implemented via `umap-learn` (version 0.5.3).[195] The embedded representations are clustered using KMeans (`n_clusters = 100`) from `scikit-learn` (version 1.8.0). While Guo et al. used seven clusters,[227] a larger number was required here to balance splits given the smaller dataset. Fewer clusters would reduce training–test similarity but exacerbate imbalance. Entire clusters are assigned exclusively to either the training or test set.

Each splitting strategy was evaluated within a three-fold CV framework. For each fold, the structural dissimilarity between training and test sets was quantified by computing Tanimoto similarities based on Morgan fingerprints. Specifically, the fingerprint of each test fragment was compared to its five nearest neighbors in the corresponding training set, and the resulting similarity distributions are shown in Figure 5.11. Among the evaluated strategies, the Butina-based split yielded the lowest similarity values, indicating the greatest structural separation between training and test sets, and was therefore selected for all subsequent CV experiments.

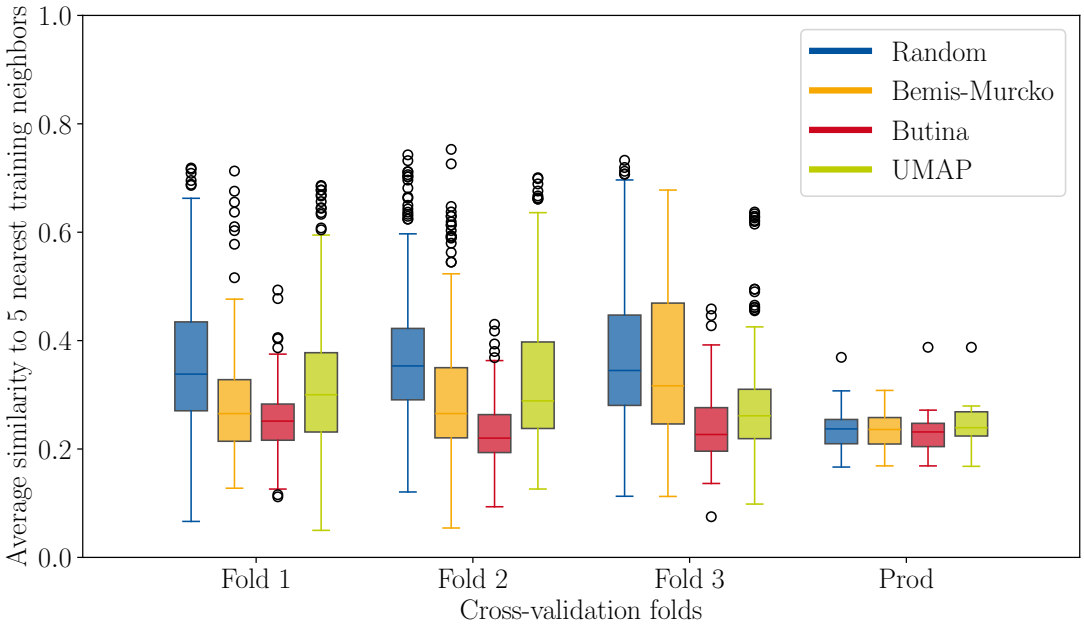


Figure 5.11: Tanimoto similarity distributions based on the Morgan fingerprints between test fragments and their five nearest neighbors in the training set for different data-splitting strategies. The Butina-based split shows the lowest similarity values, indicating greater structural dissimilarity between training and test sets.

Stratified splitting was implemented using `StratifiedGroupKFold` from `scikit-learn` (version 1.8.0)[228] to address the scarcity of positive samples. Rotational barrier values were discretized to ensure each fold contained balanced representations of low-, borderline-, and high-barrier fragments. Data partitioning within each fold allocated 53% for training, 13% for validation, and 34% for testing. Performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC were calculated per fold. The mean and standard deviation across the three folds provided a robust assessment of predictive performance and stability. For deployment in MolBar, the final production model was trained on the aggregated training and validation data from the CV splits, shuffled and re-partitioned using the Butina-based split. Generalization was evaluated using an independent benchmark dataset of 20 unique connected fragments (including 9 atropisomeric molecules), with no overlap in MolBar identifiers with the training or validation sets. This test set included all candidate axes from benchmark molecules, regardless of experimental data availability for low barriers. All learning curves and performance visualizations per fold are provided in Appendix A.3.

5.4.7 Results and Discussion

Table 5.5 summarizes the optimized parameters, including the learned atomic radii for carbon and bromine. It also specifies the exponent n, which determines the steepness of the repulsive potential. The global scaling parameter a_1 and the shift a_0, used for calibrating model output, are included.

The refined atomic radii show moderate deviations from the tabulated values by Pyykkö and Atsumi,[153] reflecting the effective steric volumes required for optimal classification. The model maintained consistent performance across all three CV folds, reliably distinguishing between high-barrier atropisomeric fragments and low-barrier non-atropisomeric fragments. Average performance metrics, including standard deviations, are summarized in Table 5.6.

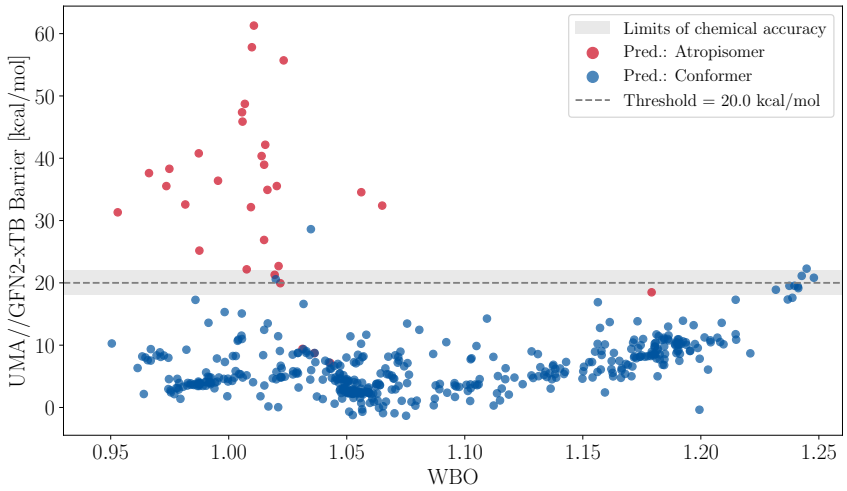
Table 5.5: Final parameters of the steric hindrance classifier. Initial radii from Pyykkö and Atsumi[153] are given in parentheses for comparison.

Parameter	Value
Carbon radius r_{C}	0.8917 Å(0.75 Å)
Bromine radius r_{Br}	0.9998 Å(1.13 Å)
Exponent n	2.8363
Global scaling a_1	3.1676
Shift a_0	-5.8445

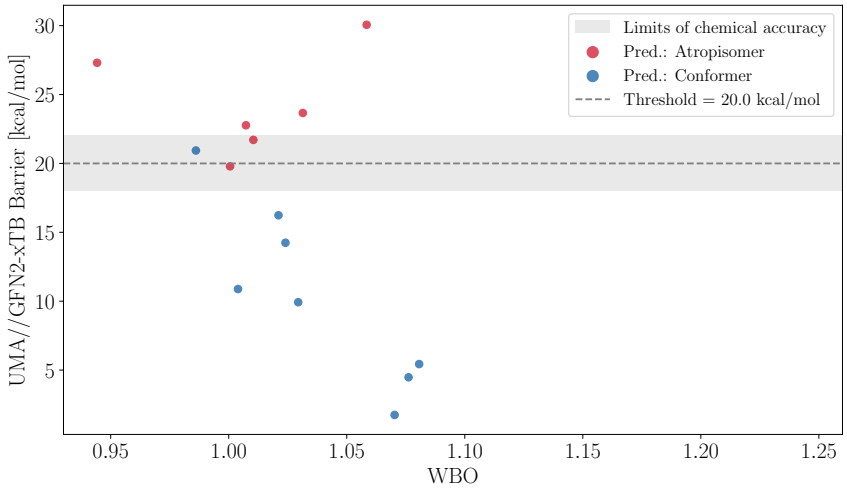
Despite the pronounced class imbalance, with only 5.9% of fragments labeled as atropisomeric, the classifier achieves an averaged ROC-AUC of $(99.74 \pm 0.10)\%$ during CV. This indicates a strong ability to rank atropisomeric fragments above non-atropisomeric ones, independent of class prevalence. The small standard deviations in accuracy and ROC-AUC underscore the model's robustness across data partitions. Precision and recall are both high and balanced, with mean values of $(88 \pm 4)\%$ and $(91 \pm 7)\%$, respectively. This balance suggests well-calibrated predictions rather than a bias toward the majority class. Misclassified fragments, detailed in Appendix A.3, are predominantly borderline cases near the 20 kcal/mol threshold, where uncertainties in computational barrier calculations complicate classification. Figure 5.12 shows the production model's predictions for both training and independent benchmark datasets. The classifier correctly identifies high-barrier compounds such as MRTX1719 (**7**, 30.0 kcal/mol) and lower-barrier examples like Elagolix (**6**, 23.7 kcal/mol) and BMS-207940 (**11**, 22.7 kcal/mol). On the benchmark dataset, the model achieves an F1-score of 87.5%, with two misclassifications: compound **10** (20.9 kcal/mol) and compound **5** (19.8 kcal/mol), both near the classification threshold.

Table 5.6: Averaged performance metrics for stratified Butina-based three-fold CV and production model performance on the independent benchmark test set. For CV, mean values and standard deviations across folds are reported.

Metric	Averaged 3-fold CV	Benchmark set
Accuracy	$(98.8 \pm 0.5)\%$	89.5%
ROC-AUC	$(99.74 \pm 0.10)\%$	96.6%
Precision	$(88 \pm 4)\%$	87.5%
Recall	$(91 \pm 7)\%$	87.5%
F1-score	$(89 \pm 4)\%$	87.5%



(a) Training set



(b) Benchmark set

Figure 5.12: Computed rotational barriers of MolBar fragments plotted against the WBO of their central bonds at the GFN2-xTB[46] level of theory. The dashed line indicates the 20 kcal/mol decision boundary employed by the steric hindrance classifier. Predictions are color-coded: red for atropisomeric fragments and blue for non-atropisomeric ones. (a) Results on the training set demonstrate reliable separation, accurately identifying high-barrier fragments as atropisomeric and low-barrier fragments as non-atropisomeric. A few misclassifications (both false positives and false negatives) appear exclusively in borderline cases near the 20 kcal/mol threshold. (b) Results on the independent benchmark set, comprising 20 unique connected fragments (including 9 genuine atropisomers). The model generalizes effectively, achieving an F1-score of 87.5%, with errors restricted to merely two cases near the boundary limit. Overall, the strong accuracy and recall emphasize the model's remarkable capability to robustly distinguish sterically hindered atropisomeric axes from freely rotating bonds. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

The consistency between CV and benchmark performance demonstrates the model's robust generalization. Residual errors occur primarily in chemically ambiguous cases near the decision boundary. This model offers a fast and reliable approach for identifying sterically hindered bonds. The results presented here establish its utility for integration into MolBar. Further application to known atropisomeric compounds, including explicit examples and MolBar

strings, is discussed in Section 6.2.

5.5 Topography Matrix: Structure Unification

5.5.1 Introduction

The rigidity matrix introduced in Equation 4.7 serves to partition a molecule into rigid fragments. These fragments allow for a conformation-independent description of the molecule's 3D structure. By definition, the conformational degrees of freedom within each rigid fragment are heavily restricted. Nevertheless, slight variations in internal coordinates can still occur among identical rigid fragments, depending on the origin of the structural data. Figure 5.13 illustrates this by depicting two minimum structures (minimum 1 and minimum 2) for the same fragment, derived from distinct computational methods (represented by black crosses on the orange and red PES). Such discrepancies may result, for instance, from the application of different DFT functionals.[147,220,229] Because the topography matrix (Equation 4.4) incorporates interatomic distances, all variations in internal coordinates must be removed to establish a reliable, consistent identifier across different conformers of a given molecule. To achieve this, a standard reference computational method is defined, and all fragments are subsequently re-optimized using this selected approach. This process of achieving structural consistency is referred to as "structure unification" or "structure idealization".

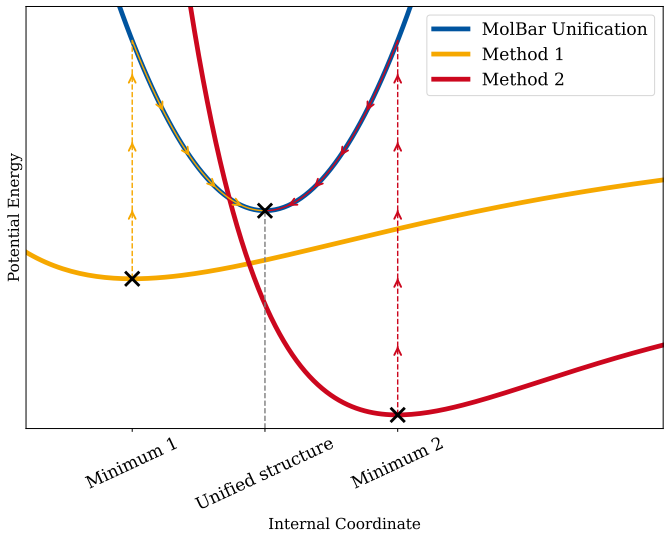


Figure 5.13: Schematic representation of the structure unification process. Two different computational methods produced two distinct minimum structures for the same fragment, minimum 1 and minimum 2 (black crosses on orange and red PES). Both fragments show variations in internal coordinates, such as bond lengths, angles, and dihedral angles. However, due to the rigidity of the fragment, both 3D structures are highly similar. Consequently, these structures occupy the same basin of attraction on the blue reference PES. The vertical dashed arrows show the projection of each minimum structure onto the unification potential surface, along which optimization then proceeds (horizontal arrows). The unification potential (blue PES) then guides the optimization along the indicated paths toward a common "unified" or "idealized" structure (black cross on blue PES).

First, a given input structure (black crosses on orange and red PES) of a fragment is projected onto the reference method's potential energy surface (blue PES), as indicated by the vertical

dashed arrows in Figure 5.13. Then, optimization is performed along the unification potential to yield the unified structure (black cross on blue PES). By definition, rigid fragments remain highly similar regardless of their source. Due to this similarity, the structures occupy the same basin of attraction on the reference method's potential energy surface. When these structures are optimized on the PES of the reference method, they converge to the same local minimum.

In geometry optimization, the objective is to identify the local minimum on the PES that corresponds to a stable molecular structure. This is typically accomplished by approximating the PES in the vicinity of the current geometry using a Taylor expansion. The simplest approach relies on a first-order (linear) expansion of the energy E around a reference geometry $\mathbf{R}_0$:

$$E(\mathbf{R}) = E(\mathbf{R}_0) + \mathbf{g}_0^{\mathrm{T}} \Delta\mathbf{R}, \tag{5.40}$$

where $\mathbf{g}_0$ is the gradient of the energy at the reference geometry, and $\Delta\mathbf{R} = \mathbf{R} - \mathbf{R}_0$ is the displacement from the reference geometry. The negative of the gradient $\mathbf{g}_0$ corresponds to the vector of forces acting on the nuclei. This vector indicates the direction of steepest descent on the PES. One can iteratively follow this steepest descent direction to reach the local minimum. A step in this direction is expressed as:[73]

$$\Delta\mathbf{R} = -\xi\mathbf{g}_0, \tag{5.41}$$

where ξ represents a step size parameter that controls the magnitude of the step taken in the direction of the negative gradient. In practical applications, the steepest descent method often exhibits slow convergence. This is particularly evident near the minimum, where the gradient diminishes in magnitude. Greater efficiency is achieved with (quasi-)Newton methods. These methods incorporate second derivative information by employing a quadratic (second-order) expansion of the energy:[73]

$$E(\mathbf{R}) = E(\mathbf{R}_0) + \mathbf{g}_0^{\mathrm{T}} \Delta\mathbf{R} + \frac{1}{2}\Delta\mathbf{R}^{\mathrm{T}}\mathbf{H}_0\Delta\mathbf{R}, \tag{5.42}$$

where $\mathbf{H}_0$ is the Hessian matrix. Differentiating the quadratic approximation in Equation 5.42 with respect to the coordinates $\mathbf{R}$ provides an approximation for the gradient at a displaced geometry:

$$\mathbf{g} = \mathbf{g}_0 + \mathbf{H}_0\Delta\mathbf{R}, \tag{5.43}$$

Since the goal is to find the minimum structure where the gradient vanishes ($\mathbf{g} = 0$), setting Equation 5.43 to zero gives:

$$\Delta\mathbf{R} = -\mathbf{H}_0^{-1}\mathbf{g}_0, \tag{5.44}$$

known as the Newton or Newton-Raphson step. When the exact Hessian is used with Newton's method, it converges to the local minimum once the PES is adequately approximated by the quadratic expansion in Equation 5.42. However, this ideal scenario is not usually the case, as the PES is more complex, so that several iterations are typically required to reach the min-

imum. If the exact Hessian is not available, quasi-Newton methods approximate the Hessian in each step using the Hessian from the previous step, the current gradient, and the geometry displacement from the previous step. One example is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method.[73] A detailed discussion of quasi-Newton methods is omitted here, because the chosen reference method (see Section 5.5.2) allows the exact Hessian to be computed directly and efficiently. Alternative approaches without the need for the Hessian are conjugate gradient methods[73] and the geometry direct inversion in the iterative subspace (GDIIS) method.[230–232]

The quadratic approximation underlying the Newton step (Equation 5.44) is only valid within a restricted region around the reference geometry. Consequently, steps extending beyond this vicinity can lead to divergence. To ensure robust convergence, step size control techniques are applied. For instance, a line search can be performed along the calculated search direction to determine a step size that sufficiently minimizes the energy. To mathematically guarantee convergence, the chosen step size must satisfy the Wolfe conditions:[73]

$$\Delta E < \alpha \mathbf{g}_0^{\mathrm{T}} \Delta \mathbf{R} \tag{5.45}$$

$$\mathbf{g}^{\mathrm{T}} \Delta \mathbf{R} > \beta \mathbf{g}_0^{\mathrm{T}} \Delta \mathbf{R}. \tag{5.46}$$

The first requirement (Equation 5.45), known as the Armijo rule, ensures a sufficient decrease in energy. The second requirement (Equation 5.46), the curvature condition, prevents excessively small steps by guaranteeing that the slope has been sufficiently reduced. A straightforward strategy to fulfill these conditions involves fitting a polynomial to the energies and gradients of both the current and previous geometries. The analytical minimum of this polynomial is then used to predict the optimal step size. Alternatively, a backtracking line search can be employed, where an initial step size is iteratively scaled down by a constant factor (e.g., halved) until the Wolfe conditions are met. Using a trust radius τ restricts the step size to a predefined limit, ensuring that the optimization remains within the valid region of the quadratic approximation. When a proposed step exceeds this radius, it is scaled down to remain within the trust radius.

The rational function optimization (RFO) method, on the other hand, employs a rational function approximation to manage the step size.[73,233] This is in contrast to the quadratic expansion in Equation 5.42. The rational function approximates the energy as follows:

$$E(\mathbf{R}) = E(\mathbf{R}_0) + \frac{\mathbf{g}_0^{\mathrm{T}} \Delta \mathbf{R} + \frac{1}{2} \Delta \mathbf{R}^{\mathrm{T}} \mathbf{H}_0 \Delta \mathbf{R}}{1 + \lambda \Delta \mathbf{R}^{\mathrm{T}} \Delta \mathbf{R}}, \tag{5.47}$$

where λ is a parameter that controls the step size. The difference between a Newton-Raphson step and an RFO step lies in the denominator of Equation 5.47. This denominator effectively limits the step size based on the magnitude of $\Delta \mathbf{R}$. It serves as a damping factor that prevents excessively large steps that could lead to divergence. Minimizing ΔE with respect to $\Delta \mathbf{R}$ by

setting $\frac{\partial \Delta E}{\partial \Delta \mathbf{R}} = 0$ leads to the RFO eigenvalue problem:[233,234]

$$\begin{pmatrix} \mathbf{H} & \mathbf{g} \\ \mathbf{g}^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} \Delta \mathbf{R} \\ 1 \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{I} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \Delta \mathbf{R} \\ 1 \end{pmatrix}, \tag{5.48}$$

Solving this generalized eigenvalue problem provides the RFO step $\Delta \mathbf{R}$. The lowest eigenvalue determines the optimal step size, balancing energy decrease and step length, while the corresponding eigenvector defines the step direction. $\Delta \mathbf{R}$ derived from Equation 5.48 is scaled down whenever it exceeds τ.[73]

Beyond step size control, the efficiency of an optimization heavily relies on the chosen coordinate system. While Cartesian coordinates $\mathbf{R}$ are a natural choice for performing the initial gradient and Hessian calculations, they often suffer from significant inter-coordinate coupling. This coupling manifests as substantial off-diagonal elements within the Hessian matrix and introduces strong anharmonicity on the PES, which can impede efficient convergence. Furthermore, analyzing the local curvature through the Hessian's eigenvalues reveals ill-conditioning, which is characterized by the presence of both very large and very small eigenvalues. To decouple these modes and alleviate these issues, the optimization can be transformed into a more chemically intuitive framework using internal coordinates $\mathbf{Q}$ (such as bond lengths, angles, and dihedrals). Adopting internal coordinates minimizes coupling and improves overall convergence stability. Conversion between Cartesian and internal coordinates is achieved using the Wilson B-matrix $\mathbf{B}$:[235]

$$\mathbf{B} = \frac{\partial \mathbf{Q}}{\partial \mathbf{R}}. \tag{5.49}$$

The Wilson B-matrix facilitates the transformation of gradients and Hessians between Cartesian and internal coordinate systems. It captures how changes in Cartesian coordinates affect internal coordinates. The gradient in internal coordinates $\mathbf{g}_Q$ is obtained from the Cartesian gradient $\mathbf{g}$ as follows:

$$\mathbf{g}_Q = \mathbf{B}\mathbf{g}. \tag{5.50}$$

The Hessian in internal coordinates $\mathbf{H}_Q$ is derived from the Cartesian Hessian $\mathbf{H}$ using:

$$\mathbf{H}_Q = \mathbf{B}\mathbf{H}\mathbf{B}^{\mathrm{T}} \tag{5.51}$$

The combination of the aforementioned Newton-Raphson steps, trust radius management, and line search techniques while using internal coordinates can result in robust and efficient geometry optimization. An example of such an approach is the ANCOpt algorithm,[234] which integrates approximate internal coordinates Q, an approximate Lindh-Hessian[236] with BFGS updates and an RFO step with trust radius management. Instead of constructing the Wilson B-matrix from a list of internal coordinates, approximate internal coordinates are derived from the eigenvectors of the Lindh-Hessian. This algorithm has been implemented in widely used semiempirical software packages such as `xtb`[70] and `CREST`.[234] The method enables efficient

geometry optimizations with minimal energy and gradient evaluations.[234]

5.5.2 Structure Unification Procedure

The primary criterion for selecting a reference method for structure unification is computational efficiency. The chosen method does not need to result in physically meaningful structures, as its sole purpose is to generate a consistent identifier. The unified structure must only preserve the connectivity and stereochemistry from the input structures. In essence, the goal is to map the varying values of bond lengths, angles, and dihedral angles onto a common reference value. This can be accomplished by employing a harmonic potential to guide the optimization toward these reference values. For example, using harmonic terms in the form of $\kappa(r_{ij}^{\text{cov}} - r_{ij})^2$ for bonds with sufficiently large force constants can force the system to the reference bond length r_{ij}^{cov}. Harmonic potentials are computationally inexpensive to evaluate, as they only require the calculation of simple quadratic terms. For all internal coordinates, the unification force field is defined as follows:

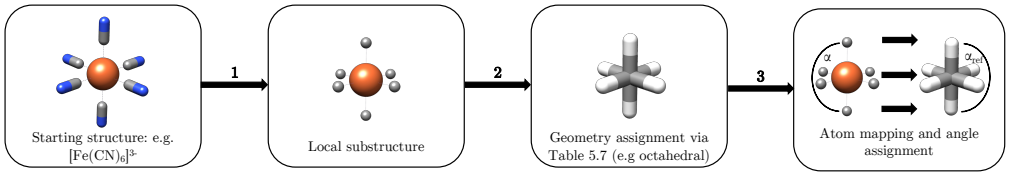
$$\begin{aligned} E_{\text{FF}} =& E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{Coulomb}} \\ =& \sum \kappa_{\text{bond}} \left(r_{ij}^{\text{cov}} - r_{ij}\right)^2 + \sum \kappa_{\text{angle}} \left(\alpha_{ijk}^0 - \alpha_{ijk}\right)^2 \\ &+ \sum \kappa_{\text{dihedral}} \Bigg(\left(\sin\theta_{ijkl}^0 - \sin\theta_{ijkl}\right)^2 \\ &+ \left(\cos\theta_{ijkl}^0 - \cos\theta_{ijkl}\right)^2 \Bigg) + \sum \frac{q_i \cdot q_j}{r_{ij}}. \end{aligned} \tag{5.52}$$

The term E_{bond} represents the sum of harmonic contributions for each bond $(i, j) \in E(G)$. Its purpose is to constrain the actual bond length to an element-pair specific reference value r_{ij}^{cov}. This reference value is determined as the sum of the covalent radii proposed by Pyykkö and Atsumi.[153] The bonds are obtained from the molecular graph $E(G)$, as detailed in Section 5.2.

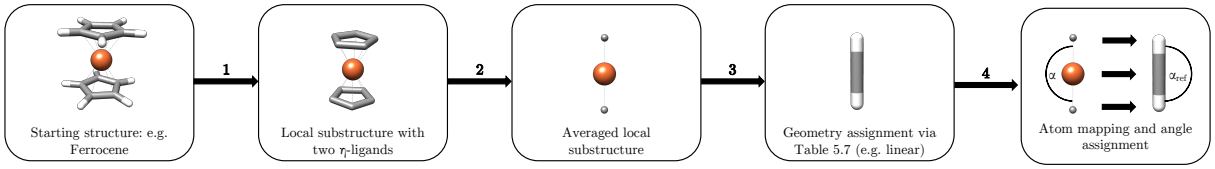
The angular energy term, E_{angle}, comprises a sum of harmonic potentials for all angles defined around each fragment core atom i. The ideal reference angles, α^0, are derived from valence shell electron pair repulsion (VSEPR) theory (see Figure 5.14).[237] In Figure 5.14a and b, step 1, for a given central atom i coordinated by neighboring atoms, the procedure begins by isolating this local substructure from the rest of the molecule. To eliminate bias from differing bond lengths during classification, all adjacent bonds are temporarily normalized to a distance of 1.5 Å.

The presence of η-ligands warrants special attention, because the individual atoms within such bound conjugated systems do not act as localized, distinct coordination sites. Instead, the entire η-ligand is handled as a single, unified coordination group (see step 2 in Figure 5.14b). To identify these ligands, the local coordination environment is analyzed by detecting interconnected substructures that exhibit elevated bond orders exceeding 1.25 (as detailed in

Section 5.3). A BFS algorithm is employed to traverse the molecular graph and systematically locate these substructures. Subsequently, the spatial coordinates of all atoms within the η-ligand are averaged to establish a single centroid, which is then used for angle assignment. For instance, the iron center in ferrocene is assigned an effective coordination number of two, which results in a linear geometry assignment determined by the centroids of its two cyclopentadienyl rings. This simplification is chemically sound, as η-ligands generally feature low rotational barriers (typically under the 20 kcal/mol MolBar threshold), allowing for nearly unrestricted rotation around the metal center.[238,239] Therefore, assigning static reference angles to individual ring atoms would be physically unrealistic. Although instances with rotational barriers exceeding 20 kcal/mol exist, they are excluded from the current framework to maintain simplicity. The only exception to this averaging procedure occurs when η-ligands are bound to other ligands, forming a bridged framework that inherently restricts rotation. In those specific scenarios, every atom of the η-ligand is considered an independent coordination group.



(a) Angle assignment algorithm without η-ligands.



(b) Angle assignment algorithm with η-ligands.

Figure 5.14: Flowchart to determine the best set of reference angles $\{\alpha^0_{ijk}\}$. First, the local structure around each atom is isolated and all bonds are set to 1.5 Å. Then, the best matching ideal VSEPR geometry is identified using the classification methods summarized in Table 5.7. When all neighboring atoms are not equivalent in the reference geometry, the Kabsch algorithm is applied to find the optimal mapping between the real local structure and the ideal geometry. For fully symmetric cases such as linear, trigonal planar or tetrahedral, the mapping is trivial. When isolated η-ligands are present, they are treated as single coordinating groups and their positions are averaged before assignment.

To illustrate the effects of this averaging scheme, consider the transition metal complex featuring an η^6-benzene ligand shown in Figure 5.15. Fragmentation divides the complex into two separate fragments: one encompassing the ring structure along with its substituents, and another comprising the metal center and its pseudo-tetrahedral coordination environment. By applying the procedure detailed in Figure 5.14b, the positions of all six benzene carbon atoms are treated as equivalent within the metal-containing fragment, as angles are assigned based on the ring's overall centroid rather than individual atomic coordinates. This approach yields a uniform representation of the η-ligand that is entirely independent of its rotational orientation around the metal. Importantly, no stereochemical information is lost during this simplifica-

tion. The relative spatial arrangement of the ligands in the pseudo-tetrahedral environment is explicitly preserved, and the isolated ring fragment fully retains its inherent planar chirality.

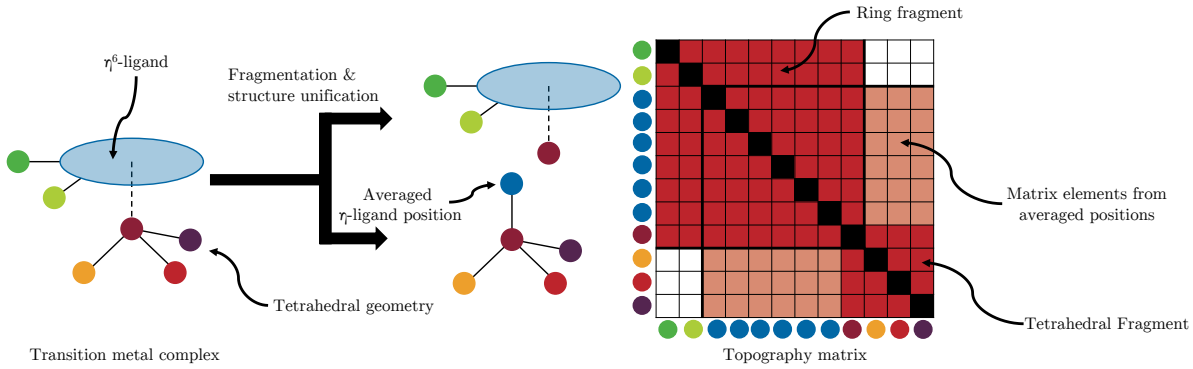


Figure 5.15: Illustration of the positional averaging scheme applied to a transition metal complex featuring an η^6-benzene ligand. During fragmentation, the complex is partitioned into two distinct fragments: the ring substructure (comprising the six carbon atoms, their substituents, and the metal center) and the pseudo-tetrahedral metal substructure (comprising the metal center and the non-ring substituents). By treating the η^6-benzene ring as a single coordinating entity anchored at its centroid, the structural influence between the ring and the non-ring ligands becomes uniform for all six carbon atoms. This approach preserves all essential stereochemical information, including the relative spatial arrangement around the metal center and the planar chirality of the substituted ring. Crucially, it guarantees that the generated topography matrix is strictly invariant to the free rotation of the η-ligand.

The best-matching ideal VSEPR geometry is determined using the empirical classification methods outlined in Table 5.7. These rules have been specifically designed and optimized to ensure robust geometry assignments across diverse molecular topologies. The classification approach systematically integrates coordination number, spatial volume, and bond angle analyses. Following this initial geometric evaluation, the algorithm applies a set of element-specific override rules to enforce essential chemically informed constraints. Crucially, all second-period elements ($Z \leq 10$) and all transition metals with an effective CN of 3 are strictly reassigned to a trigonal planar geometry. For atoms such as nitrogen, which naturally exhibit a locally trigonal pyramidal structure, this reassignment correctly accounts for their low geometric inversion barriers. Because pyramidal nitrogen rapidly interconverts at room temperature, the time-averaged effective geometry is trigonal planar; stereochemically, such centers are inherently achiral under standard conditions. Assigning a planar reference geometry captures this achirality. However, formal chiral nitrogen centers occasionally exist within strained ring systems where the macroscopic ring framework restricts the inversion process, trapping the nitrogen in a stable pyramidal conformation. In these specialized cases, the high structural tension of the surrounding ring dominates the potential energy surface. Consequently, even though the algorithm formally applies a planar angle constraint to the isolated atom, the rigid geometric demands of the neighboring rigid fragments force the assembled molecule to remain pyrami-

Table 5.7: VSEPR geometry classification methods and assignment criteria for different coordination numbers.

CN	Geometries	Classification Method
0	dot	Trivial assignment for isolated atoms
1	linear CN=1	Trivial assignment for terminal atoms
2	linear CN=2, bent	Bond angle analysis: angles $> 160°$ favor linear, smaller angles favor bent. Override: non-ring atoms without any elevated WBO assigned linear.
3	trigonal planar, trigonal pyramidal	Volume analysis: $V = \lvert\det(\mathbf{R})\rvert$ where $\mathbf{R}$ is 3×3 coordinate matrix. Near-zero volume (< 0.5 Å^3) indicates planar, larger volumes indicate pyramidal. Override: elements with Z ≤ 10 and transition metals are assigned planar.
4	tetrahedral, square planar	Planarity deviation analysis: average distance $\bar{d}$ of atoms from all $\binom{4}{3} = 4$ possible three-atom planes. High deviation (> 1.5 Å) indicates 3D tetrahedral, low deviation indicates planar arrangement. Override: C atoms assigned tetrahedral.
5	trigonal bipyramidal, square pyramidal	Plane analysis: evaluate all $\binom{5}{3} = 10$ three-atom plane configurations. Score based on remaining atom distribution: 0 additional atoms in-plane (threshold 0.3 Å) scores trigonal (three atoms in one plane, while two are perpendicular), 1 atom in-plane scores square (four atoms in one plane, one perpendicular to plane).
6	octahedral	Trivial assignment for octahedral geometry
7	pentagonal bipyramidal	Trivial assignment for pentagonal bipyramidal geometry
8	square antiprismatic	Trivial assignment for square antiprismatic geometry
> 8	Coulomb	Fallback assignment for unsupported coordination numbers, being only constrained by Coulomb repulsion to maximize interatomic distances.

dal in its unified state. Thus, true macrocyclic stereochemistry is robustly preserved without requiring complex local exceptions to the planar nitrogen rule. Similarly, carbon atoms with a CN of 4 are universally enforced as tetrahedral. For non-ring atoms with a CN of 2, the geometry is dictated by the local bond order: atoms involved in double bonds are assigned a linear geometry, while purely single-bonded constituents are assigned a bent geometry.

After determining the ideal VSEPR geometry class for each fragment atom, the corresponding reference angles α_{ijk}^0 are obtained from tabulated VSEPR angle values. In cases where the reference geometry possesses non-equivalent coordination sites (e.g., axial versus equatorial positions in a trigonal bipyramid), the Kabsch algorithm is utilized to compute the optimal spatial mapping between the actual atomic positions and the ideal reference points.[240] This algorithm computes the optimal rotation matrix to align both structures. As it is dependent on the atomic order, all possible atomic permutations of the real local structure must be considered to find the best match. The best match according to the RMSD criterion is then used to extract the reference angles.

E_{dihedral} represents the sum of harmonic terms for all dihedral angles used to constrain certain structures within the molecule. For instance, the dihedral constraint ensures that bonds with elevated bond orders, as determined in Section 5.3, are maintained in a specific conformation. Additionally, aromatic rings are constrained to remain planar. These terms are also applied to planarize aliphatic ring structures (e.g., cyclohexane), unifying different conformations into a single structure. As previously discussed, the unified structure does not need to be physically meaningful, provided it preserves the connectivity and stereochemistry of the input. Crucially, this means retaining the relative positions of substituents—specifically, whether atoms lie above or below the ring plane. Mapping different ring conformations (e.g., chair and boat) onto a single planar structure yields a consistent representation of the ring system regardless of its conformation. Thus, the topography matrix in Equation 4.4 remains conformer-independent, even for non-aromatic ring systems, while still encoding the information required to distinguish *cis*/*trans* isomers, for example. The sine and cosine basis was selected to avoid unreasonably high energy contributions during geometry optimization, which could arise from a possible sign change of a dihedral angle near $180°$. By employing the sine and cosine basis, as shown in Equation 5.52, the sign reversal issue is circumvented, yielding a more stable representation of the dihedral angles, since $\sin(180°) = \sin(-180°) = 0$ and $\cos(180°) = \cos(-180°) = -1$.

E_{Coulomb} is introduced to enforce repulsive interatomic interactions, which results in maximizing the interatomic distances under constraints. Thus, E_{Coulomb} sums over all atom-atom combinations with element-independent partial charges q_i and q_j ($q_i \cdot q_j = \text{constant} > 0$). Table 5.8 shows the default values for the parameters used in the force field to unify the input structures.

Table 5.8: Default values for the scaling constants used in the force field to unify the input structures.

constant	default value
κ_{bond}	$1 \cdot 10^6$ Å^{-2}
κ_{angle}	$2 \cdot 10^3$ rad^{-2}
κ_{dihedral}	$2 \cdot 10^3$
$q_i \cdot q_j$	$1 \cdot 10^2$ Å

Having fully formulated the unification potential and established the individual force field terms for a given fragment, the objective is to locate the corresponding local energy minimum through geometry optimization. Because the energy expression in Equation 5.52 is composed entirely of simple harmonic and Coulombic terms, its first and second derivatives (the gradient vector and the Hessian matrix, respectively) can be computed analytically with high computational efficiency. The complete derivation of the analytical gradient and Hessian equations for this specific force field is detailed in Appendix B. To ensure the robustness of the implementation, both the analytical gradient and Hessian were rigorously verified against numerical derivatives prior to deployment.

Because the unification procedure may require optimizing a large number of fragments, the energy minimization algorithm must be both robust and computationally efficient. Otherwise, the optimization step could easily become the rate-limiting stage in the overall MolBar calculation workflow. A suitable foundation for developing such an algorithm is the ANCOpt method introduced earlier. However, preliminary investigations reveal that the original ANCOpt implementation in CREST requires modifications to address the distinct challenges presented by the unification potential. The energy, gradient, and Hessian in this context are fundamentally different from those in quantum chemical methods. The large force constants in the unification potential create steep energy landscapes with high curvatures, where the energy unit is arbitrarily scaled up compared to typical quantum chemical approaches in Hartree or eV. These features require careful consideration of numerical stability and step size control during optimization. Using CREST's ANCOpt implementation without modifications led to numerical instabilities and convergence issues. Further, the original ANCOpt algorithm can be modified to take advantage of the low computational cost of energy evaluations for the unification potential. CREST employs a Davidson solver[241] to determine the lowest eigenvalue and its corresponding eigenvector of the RFO matrix in Equation 5.48. However, for RFO step-size control, the Davidson solver proved highly inconsistent when applied to the unification potential. In certain test cases, it required an excessive number of iterations to converge, ultimately making it slower than `NumPy`'s interface to LAPACK's `syevd`,[134] while in others, it provided the anticipated acceleration. Although the `syevd` routine offers greater stability, its computational cost scales cubically with the size of the system. Given that energy evaluations for the unification potential are computationally inexpensive, repeated calculations, such as those in a backtracking line search, introduce only negligible overhead. This observation

motivates replacing the RFO step-size control with a backtracking line search approach. The latter substitutes the computationally demanding task of finding the eigenvector associated with the lowest eigenvalue of the RFO matrix with a more economical Newton step and energy evaluations, which are effectively cost-free. MolBar's unification potential further enables both computationally efficient and exact evaluations of the Hessian matrix. As a result, quasi-Newton BFGS updates of the Hessian can be replaced with exact Hessian evaluations at each optimization iteration.

The MolBar-specific optimization algorithm tailored for the unification procedure is referred to as MolBar optimizer (MolBarOpt). Similar to the ANCOpt method, MolBarOpt operates in approximate normal coordinates derived from the eigenvectors of the exact Hessian, which are reconstructed every k optimization steps. To decouple internal vibrations from overall translation and rotation, the molecular coordinates are first shifted to their geometric center. A matrix $\mathbf{P}' \in \mathbb{R}^{3n_{\text{atoms}} \times 6}$ is then constructed to capture the translational and rotational modes.[234] Each row corresponds to a Cartesian degree of freedom of an atom, while each column represents one of the six rigid-body modes. For linear molecules—identified by analyzing the eigenvalues of the inertia tensor—only five modes are projected out, as rotation around the molecular axis does not alter the geometry. Orthonormalization of $\mathbf{P}'$, for example via the Gram-Schmidt process,[242] yields the finalized projection matrix $\mathbf{P} \in \mathbb{R}^{3n_{\text{atoms}} \times n_{\text{rigid}}}$, where n_{rigid} is 5 or 6. The rigid-body modes are formally removed from the Hessian and gradient using this projection operator:

$$\mathbf{H}_{\text{proj}} = (\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})\mathbf{H}(\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}}), \tag{5.53}$$

$$\mathbf{g}_{\text{proj}} = (\mathbf{I} - \mathbf{P}\mathbf{P}^{\mathrm{T}})\mathbf{g}. \tag{5.54}$$

Following this projection, $\mathbf{H}_{\text{proj}}$ is free of translational and rotational contributions, as their corresponding eigenvalues are effectively zeroed out. To transition the optimization into internal coordinates, the approximate normal coordinate (ANC) basis is constructed by diagonalizing the projected Hessian:

$$\mathbf{H}_{\text{proj}} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}. \tag{5.55}$$

Here, $\mathbf{U}$ is an orthogonal matrix of eigenvectors representing the normal modes, and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues characterizing the local curvature along each mode. The ANC basis matrix $\mathbf{B}$ is formally defined by retaining only the physically meaningful eigenvectors:[234]

$$\mathbf{B} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{n_{\text{modes}}}], \tag{5.56}$$

where $n_{\text{modes}} = 3n_{\text{atoms}} - n_{\text{rigid}}$ is the number of retained vibrational modes. This matrix linearly relates the Cartesian coordinates to the internal normal coordinates, allowing the gradient and Hessian to be transformed accordingly:[243]

$$\mathbf{g}_{\text{internal}} = \mathbf{B}^{\mathrm{T}}\mathbf{g}_{\text{proj}}, \quad \mathbf{H}_{\text{internal}} = \mathbf{B}^{\mathrm{T}}\mathbf{H}_{\text{proj}}\mathbf{B}. \tag{5.57}$$

Now, with the gradient and Hessian expressed in internal coordinates, the optimization step can be computed. The optimal step in internal coordinates is determined using a Newton-Raphson approach combined with backtracking line search. First, the Newton direction in internal coordinates $\Delta\mathbf{Q}_{\text{Newton}}$ is computed by solving the linear system:[243]

$$\mathbf{H}_{\text{internal}}\Delta\mathbf{Q}_{\text{Newton}} = -\mathbf{g}_{\text{internal}}. \tag{5.58}$$

The Newton direction is then scaled using a backtracking line search to ensure sufficient energy decrease. Starting with an initial step size $\xi = 1$, the algorithm iteratively reduces this factor until the Armijo condition is satisfied (see Equation 5.45).[244] The final step in internal coordinates is computed as $\Delta\mathbf{Q} = \xi\Delta\mathbf{Q}_{\text{Newton}}$. If the Newton direction is not a descent direction (i.e., $\mathbf{g}_{\text{internal}}^{\text{T}}\Delta\mathbf{Q}_{\text{Newton}} \geq 0$), the algorithm falls back to a steepest descent step. The final Cartesian step is derived by transforming the internal coordinate step back to Cartesian coordinates using the ANC basis matrix:

$$\Delta\mathbf{R} = \mathbf{B}\,\Delta\mathbf{Q}. \tag{5.59}$$

This step is applied to update the molecular geometry as follows:

$$\mathbf{R}_k = \mathbf{R}_{k-1} + \Delta\mathbf{R}. \tag{5.60}$$

This procedure is repeated iteratively until the convergence criterion is satisfied. The convergence criterion evaluates the change in geometry over several iteration steps. Since the ultimate objective is to obtain a unified topography matrix, the focus is on changes in these distances rather than solely on energetics. For each iteration step, the maximum change in interatomic distances is calculated according to:

$$\Delta r_{\text{max}} = \max_{i<j} |r_{ij}^{(k)} - r_{ij}^{(k-1)}|, \tag{5.61}$$

where $r_{ij}^{(k)}$ represents the distance between atoms i and j at iteration k. The optimization is deemed converged when Δr_{max} falls below a predefined threshold (e.g., 1×10^{-2} Å) for a specified number of consecutive iterations (e.g., 5 iterations). Due to large force constants in the unification potential, using the gradient norm as a convergence criterion can be misleading. The gradient norm may remain large even when the geometry has stabilized, as small changes in geometry can lead to significant energy variations. Focusing on changes in interatomic distances directly addresses the goal of achieving a consistent geometry for the unification procedure.

In the following section, the performance of the MolBarOpt algorithm is benchmarked against the ANCOpt algorithm and the Newton-CG method from `SciPy`.

5.5.3 Dataset

The benchmark dataset is derived from the BACE collection, which comprises 1,513 drug-like molecules.[224,225] The BACE dataset consists of CREST ensembles for each molecule, which were generated at the GFN2-xTB[70] level of theory. SMILES strings were used to ensure stereochemical integrity of the provided conformers, confirming that no unintended inversions at chiral centers or double bonds occurred during the initial CREST conformational sampling. 10 conformers were randomly sampled for each molecule from the CREST ensembles, yielding an initial pool of 15,130 structures. Applying the MolBar fragmentation procedure (outlined in Sections 5.3 and 5.4) to this pool generated 206,704 individual fragments. Because common substructures, such as C_2H_3 or C_3H_2, appear repeatedly across different parent molecules, the fragments were categorized based on their kekulized SMILES representations. The SMILES was used instead of the MolBar because MolBar is being tested in this benchmark. Filtering based on the MolBar would require carrying out the optimization procedure outlined here to calculate the MolBar string, which would bias the results. To prevent structural redundancy from skewing the benchmark, the dataset was filtered to retain exactly 10 conformations per unique fragment.

The resulting curated dataset consists of 788 unique fragments, providing a final total of 7,880 structures for the benchmark analysis. These fragments span a wide range of sizes, from simple three-atom motifs, such as the C-O-H unit, to larger substructures containing up to 32 atoms. The complete distribution of these fragment sizes is depicted in Figure 5.16.

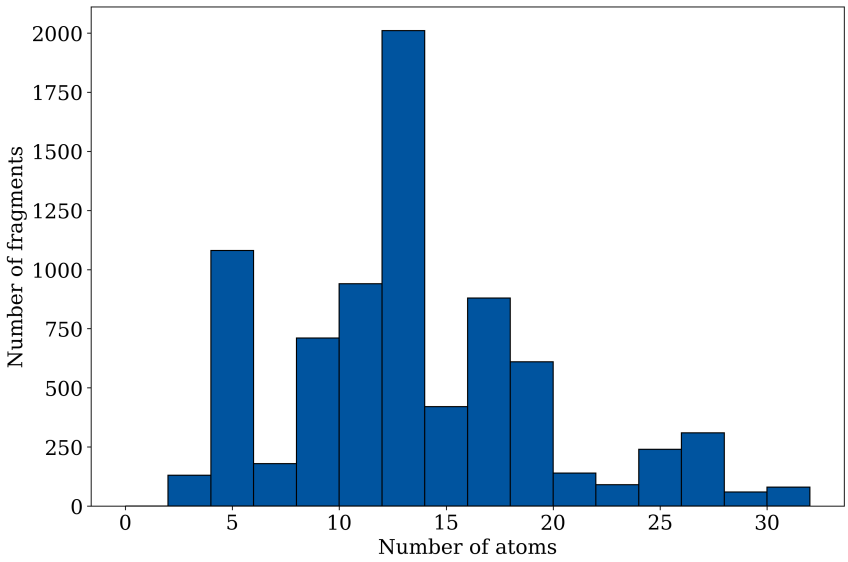


Figure 5.16: Distribution of fragment sizes in the MolBar fragment dataset derived from the BACE[224,225] dataset, encompassing 7,880 fragments with sizes ranging from 3 to 32 atoms.

The average fragment size is 13 atoms, with a standard deviation of 6 atoms. The median fragment size is 12 atoms. A dataset with larger fragments was not selected, as this distri-

bution aligns with the typical output of the MolBar fragmentation procedure for drug-like molecules. Larger fragments are relatively uncommon, while small rings, methyl groups, or functional groups occur more frequently. To include some larger systems as test cases, the peptide Microviridin,[245] comprising 225 atoms, as well as a Möbius 50-cyclacene structure[246] with 300 atoms and the large fullerene C_{720}[247] with 720 atoms, are also included.

5.5.4 Computational Details

The benchmark evaluates the performance of three distinct optimization algorithms:

- MolBarOpt: The developed optimization algorithm in this work, specifically tailored for the unification potential, as described in the preceding section.
- ANCOpt: A Python-based adaptation of the original ANCOpt algorithm from `CREST`,[234] modified to function with the unification potential.
- Newton-CG: The Newton-Conjugate Gradient method from the `SciPy` library, a widely recognized optimization algorithm.

This study employs a Python-based implementation of CREST's ANCOpt algorithm. The decision to use this implementation is driven by the fact that the original CREST code is optimized for quantum chemical potentials in Hartree units and would necessitate substantial modifications for compatibility with the unification potential. The Python-based version adheres to the original ANCOpt algorithm but incorporates adjustments tailored to the unification potential. As previously noted, the Davidson solver used in CREST for solving the RFO eigenvalue problem demonstrates inconsistent performance and convergence challenges when applied to the unification potential. Similar issues arose when using `SciPy`'s interface to the ARPACK `sseupd` eigensolver. A more efficient alternative to the full diagonalization using `NumPy`'s interface to `syevd` (`np.linalg.eigh`) is the LAPACK `dsyevr` routine, which computes selected eigenvalues and eigenvectors of a real symmetric matrix using the Relatively Robust Representations algorithm.[248] Thus, for the ANCOpt clone, a custom Python extension module for `dsyevr` is utilized to determine the lowest eigenvalue and corresponding eigenvector of the RFO matrix in Equation 5.48. Unlike CREST's Davidson solver or `sseupd`, `dsyevr` consistently converges to yield one eigenvalue/eigenvector pair in the case of the unification potential. For the fullerene C_{720}, as an example, 83% of the total wall time in this Python-based ANCOpt implementation is spent diagonalizing Hessians via `syevd` for ANC basis construction and solving the RFO eigenvalue problem using `dsyevr`. Thus, a transition to a full `Fortran` implementation, as in CREST, is not expected to yield significant changes in relative performance compared to the original implementation. Unlike the original implementation, instead of quasi-Newton BFGS updates, the exact Hessian is computed at each iteration, following the approach used in MolBarOpt. This is done because Hessian evaluations are computationally inexpensive for the unification potential.

Each fragment derived from the BACE[224,225] dataset is optimized using all three algorithms, starting from identical initial geometries. The Newton-CG implementation is taken from the `SciPy` library, version 1.16.2. The ANCOpt optimizer employs `NumPy`, version 2.3.3. In contrast to the original CREST implementation, which updates the ANC basis every 20 iterations, both MolBarOpt and ANCOpt Python-clone extend this interval to 100 iterations. Preliminary testing indicated that more frequent basis reconstructions provide no meaningful performance advantage. All optimization algorithms are allowed a maximum of 1000 iterations to achieve convergence. All three algorithms are implemented in Python and share identical force field constraints, with analytical expressions for energy, gradient, and Hessian (Equation 5.52) implemented in `Fortran`. All three algorithms utilize the same `Fortran` implementation for energy, gradient, and Hessian evaluations and the same convergence criterion based on changes in interatomic distances (see Equation 5.61). Timings are recorded as wall-clock time on an Apple M1 Pro CPU with 10 cores and 32 GB RAM, running macOS 26.1. The timings are averaged for all fragments and their conformers of the same size to analyze scaling behavior. For the special test cases of larger molecules (Microviridin, Möbius 50-cyclacene, and C_{720}), the timings were averaged over five independent runs.

5.5.5 Results and Discussion

The benchmarking results across the 7,880 fragments from the BACE dataset[224,225] establish MolBarOpt as the most efficient and robust method for structure unification (Table 5.9 and Figure 5.17).

Table 5.9: Benchmark results for the optimization of 7,880 fragments from the BACE[224,225] dataset using MolBarOpt, ANCOpt, and Newton-CG.

Metric	MolBarOpt	ANCOpt	Newton-CG
# topography errors	0	553	15
Mean Gradient Norm (Å^{-1})	1.94×10^{0}	3.02×10^{3}	4.14×10^{0}
Median Gradient Norm (Å^{-1})	1.10×10^{-8}	1.46×10^{-4}	2.73×10^{-1}
Mean Wall Time (ms)	5.85	9.14	12.22
Median Wall Time (ms)	4.84	7.29	7.36
Mean Iterations	14.3	17.2	30.8
Median Iterations	12	15	19
Mean Time/Iteration (ms)	0.38	0.47	0.35
Mean Speedup (vs Newton-CG)	1.86	1.29	1.00
Median Speedup (vs Newton-CG)	1.53	1.12	1.00
Scaling prefactor (ms)	0.06	0.05	0.05
Scaling exponent	1.72	1.94	2.09

The convergence quality of MolBarOpt is demonstrably superior. It achieves a median gradient norm of 1.10×10^{-8} Å^{-1}, which is four orders of magnitude lower than ANCOpt (1.46×10^{-4} Å^{-1}) and seven orders of magnitude lower than Newton-CG (2.73×10^{-1} Å^{-1}).

This performance advantage is clearly visible in the gradient norm distributions depicted in Figure 5.17a, where MolBarOpt consistently attains the lowest gradient norms across the entire dataset.

Because the convergence criterion relies strictly on changes in interatomic distances rather than energetic metrics, a specific, universal threshold for the final gradient norm cannot be defined. The unification potential is characterized by steep energy wells; thus, even minor geometric deviations can manifest as large gradient residuals. For highly constrained systems—where multiple geometric constraints fundamentally oppose one another—the gradient norm may remain relatively elevated even after the geometry has been thoroughly symmetrized and stabilized. Consequently, while higher gradient norms do not strictly equate to optimization failure, comparing these values between methods remains a reliable proxy for assessing which algorithm found a more stable geometry.

Crucially, MolBarOpt is the only algorithm to complete the benchmark with zero topography errors. A topography error arises when the eigenvalue spectra of the fragment topography matrices differ across the initial conformers of a given fragment, indicating a failure to adequately unify all identical structural inputs. By avoiding these errors entirely while simultaneously minimizing the final gradient norms, MolBarOpt proves it consistently maps all diverse conformers of a fragment to the exact same unified reference structure—the ultimate goal of the unification procedure. While MolBarOpt achieves excellent convergence in most instances, a small subset of fragments exhibit relatively high final gradient norms (around 10^2 Å^{-1}). Nevertheless, manual inspection of these edge cases confirms that the geometries are successfully symmetrized, corresponding to the aforementioned highly constrained systems. Because no topography errors occur for these cases, the elevated gradient norms do not indicate an underlying convergence failure.

The Newton-CG method also demonstrates a high level of robustness, with only 15 fragments (0.19%) producing topography errors. In contrast, while ANCOpt achieves final gradient norms comparable to MolBarOpt for many systems, it also produces a significant number of cases with extremely high gradient norms (around 10^3 Å^{-1}), indicating severe convergence issues. This dichotomy is further supported by the substantial deviation between the mean (3.02×10^3 Å^{-1}) and median (1.46×10^{-4} Å^{-1}) gradient norms for ANCOpt. Such poor convergence directly translates to the 553 topography errors (7.02% of all fragments) observed for ANCOpt, showing that different initial geometries of the same fragment frequently map to different unified structures. Overall, because the primary algorithmic difference between MolBarOpt and ANCOpt lies in the step-size control mechanism (backtracking line search vs. RFO), the improved convergence robustness of MolBarOpt can be directly attributed to its more reliable step-size selection.

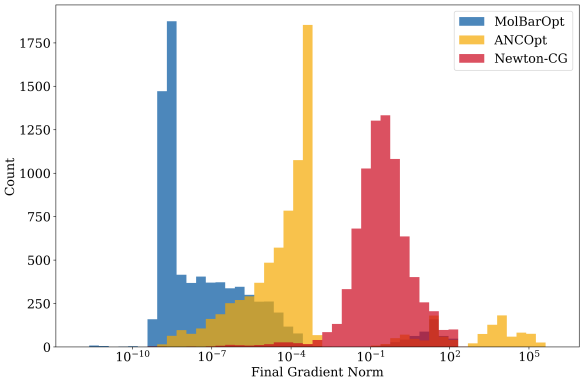
In terms of computational efficiency, MolBarOpt delivers superior performance compared to its competitors, achieving a mean speedup of $1.86\times$ relative to Newton-CG. ANCOpt, by contrast, attains a mean speedup of only $1.29\times$. On average, MolBarOpt requires 5.85 ms to

optimize a fragment, whereas ANCOpt takes 9.14 ms and Newton-CG requires 12.22 ms. This distinct performance advantage is illustrated in the wall time distributions presented in Figure 5.17b. The histogram reveals three rough clusters of wall times corresponding to size: small fragments (3–6 atoms, ∼1 ms), medium fragments (7–15 atoms, ∼10 ms), and large fragments (16–32 atoms, below 100 ms). However, computational cost is heavily influenced by the intrinsic flexibility of the fragment. For instance, aliphatic rings increase wall times because they must be forcibly planarized during the unification procedure, requiring significant deviation from their initial geometries. Conversely, aromatic rings are already planar and demand less optimization effort. Consequently, these size-based clusters in the wall time distributions lack sharply defined boundaries.

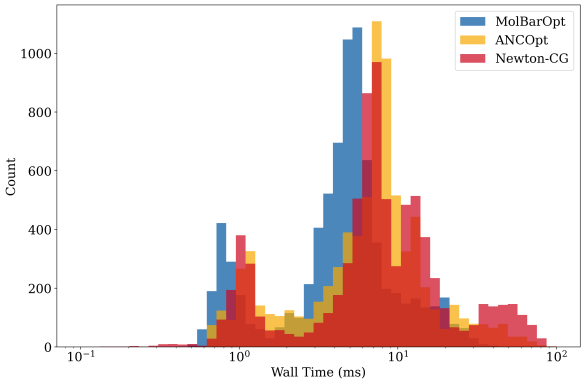
Finally, MolBarOpt achieves convergence with the fewest iterations, averaging 14.3 iterations (0.38 ms per iteration), compared to ANCOpt (17.2 iterations, 0.47 ms per iteration) and Newton-CG (30.8 iterations, 0.35 ms per iteration). Although Newton-CG and MolBarOpt exhibit similar wall times per iteration, the overall computational cost of MolBarOpt (5.85 ms) remains significantly lower than that of Newton-CG (12.22 ms), primarily due to the drastically reduced number of iterations required to reach convergence.

Overall, MolBarOpt demonstrates superior performance in optimization quality but also in computational efficiency compared to the other two algorithms. The scaling behavior of the three algorithms with fragment size is further analyzed in Figure 5.17c. The scaling exponents were determined by fitting a power law of the form $t = a \cdot N^b$ to the averaged wall time per number of atoms. MolBarOpt exhibits the most favorable scaling behavior, with a scaling exponent of 1.72, compared to 1.94 for ANCOpt and 2.09 for Newton-CG. The prefactors a for each method were also extracted from the fits, which are similar for all methods (0.05–0.06 ms). The improved scaling of both MolBarOpt and ANCOpt can be attributed to the use of the ANC basis, which decouples vibrational modes and reduces the number of iterations required for convergence. Unnecessary steps, such as those affecting only translational or rotational modes without altering the internal geometry, are avoided. This decoupling is especially beneficial for larger systems, where Cartesian coordinates often lead to inefficient optimization paths due to mode coupling. When comparing the two ANC-based optimizers, the enhanced scaling of MolBarOpt can be attributed to its backtracking line search mechanism for step-size control. The line search ensures efficient step sizes without the need to solve the RFO matrix eigenvalue problem, which becomes computationally expensive for larger systems. Given the low cost of energy evaluations for the unification potential, the backtracking line search introduces minimal overhead compared to the eigenvalue problem in RFO.

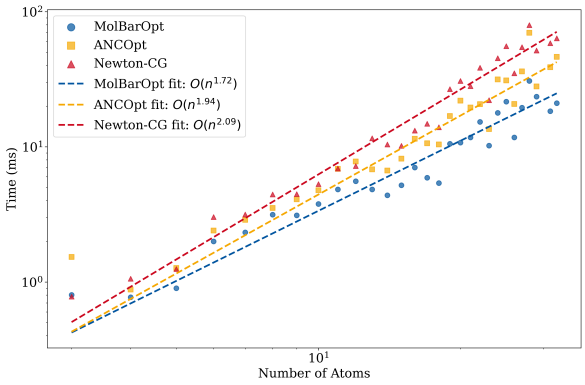
To analyze the performance on more challenging systems, three representative case studies are examined in detail: the peptide Microviridin[245] (225 atoms), a Möbius 50-cyclacene structure[246] (300 atoms), and the buckminster fullerene C_{720}[247] (720 atoms). Microviridin would typically be fragmented into smaller, rigid substructures by MolBar. However, to evaluate the optimizer on larger, highly flexible, and more demanding cases, the full peptide is considered here. In



(a) Final Gradient Norm



(b) Wall Time Distribution



(c) Computational Cost Scaling

Figure 5.17: Benchmarking results of three optimization algorithms on the BACE fragment dataset[224,225] for the MolBar unification potential. (a) The final gradient norms achieved by each algorithm are compared, with MolBarOpt (blue) achieving the lowest mean and median gradient norms, indicating superior optimization quality. (b) The wall time distribution for each algorithm is presented, showing that MolBarOpt generally requires less time to converge compared to ANCOpt (orange) and Newton-CG (red). (c) The computational cost scaling with fragment size is illustrated, revealing that MolBarOpt exhibits the most favorable scaling behavior, maintaining lower wall times as fragment size increases. Overall, MolBarOpt demonstrates superior performance in terms of optimization quality, speed, and scalability compared to the other two algorithms.

contrast, both the Möbius structure and the C_{720} fullerene represent rigid systems that are treated as single fragments in MolBar.

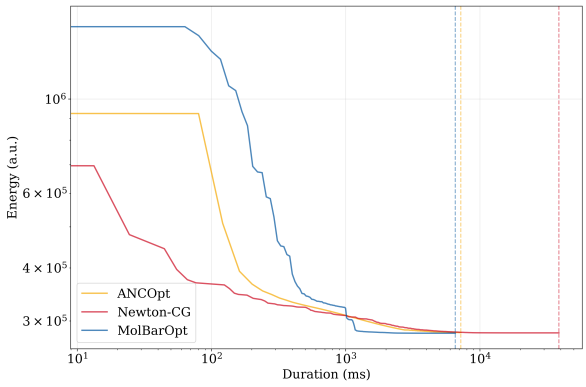
For the Microviridin structure (Figure 5.18), MolBarOpt demonstrates clearly superior convergence properties, terminating with a final gradient norm of 4.38×10^{-4} Å^{-1} in 6.57 seconds. The geometric idealization of this peptide primarily entails maximizing interatomic distances through Coulombic repulsion, while strictly enforcing local planarity and VSEPR angular constraints (Figure 5.18c and d). In contrast, ANCOpt terminates after 7.22 seconds with an elevated gradient norm of 4.07×10^{3}. This indicates premature convergence, suggesting the algorithm becomes trapped in a region of the PES characterized by steep gradients, thereby failing to locate the true local minimum. This comparison emphasizes the efficacy of the backtracking line search implemented in MolBarOpt; it maintains sufficient flexibility to navigate complex energy landscapes and efficiently locate deep minima where the RFO approach struggles. The kinetic behavior of the optimizations is further detailed in the gradient norm convergence plot shown in Figure 5.18b. Once the optimization trajectory enters the near-quadratic region of the PES, the pure Newton steps employed by MolBarOpt induce a rapid, asymptotic reduction in the gradient norm. Conversely, ANCOpt fails to progress further, severely hampered by sub-optimal step-size estimations generated by the RFO method. The Newton-CG method, operating purely in Cartesian coordinates, requires 38.77 seconds, more than five times the duration of MolBarOpt. However, unlike ANCOpt, Newton-CG successfully navigates the landscape to reach a final gradient norm of 5.58 Å^{-1}, confirming convergence to the local minimum. This disparity highlights the fundamental advantage of the ANC basis utilized by both MolBarOpt and ANCOpt: by effectively decoupling the internal vibrational modes from translation and rotation, the ANC framework significantly reduces the required optimization steps, allowing both methods to drastically outperform the Cartesian-based Newton-CG approach in per-iteration efficiency.

The Möbius 50-cyclacene shows similar trends (Figure 5.19). Here, the structure idealization again makes the ring more circular, while maintaining planarity of local structures of the benzene rings (Figure 5.19c and d). MolBarOpt converges in 4.66 seconds with a final gradient norm of 9.39×10^{-2} Å^{-1}, while ANCOpt takes longer (7.13 seconds) and achieves a higher final gradient norm of 3.64 Å^{-1}. In this case, Newton-CG requires less time than ANCOpt at 5.59 seconds with a final gradient norm of 6.19×10^{-1} Å^{-1}. This indicates that for such a large system solving the RFO eigenvalue problem becomes more expensive, and the benefits of the ANC basis are less pronounced due to the rigidity of the structure. However, MolBarOpt still achieves the best convergence quality with the lowest gradient norm.

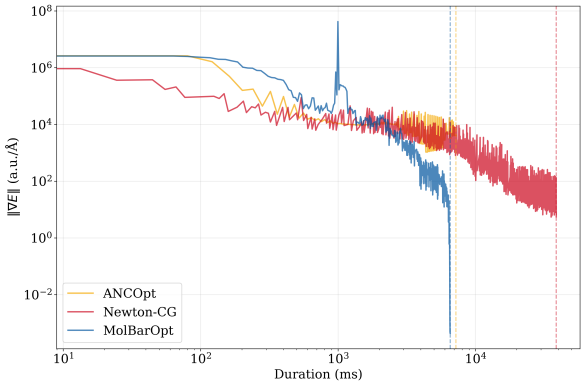
The C_{720} fullerene (Figure 5.20) reveals an interesting crossover in algorithm performance. The structure unification leads to a more aspherical shape instead of the nearly perfect spherical geometry of the initial structure (Figure 5.20c and d). This highlights the chiral nature of the molecule. For this 720-atom system, Newton-CG achieves the fastest convergence at 4.21 seconds, while MolBarOpt requires 27.15 seconds and ANCOpt 61.71 seconds. This reversal

occurs because the overhead of constructing and diagonalizing the Hessian to build the ANC basis becomes significant for very large systems. For C_{720}, the Hessian is a 2160×2160 matrix, and its diagonalization dominates the computational cost. Additionally, due to the rigidity of the fullerene structure, the benefits of the ANC basis in decoupling vibrational modes are less pronounced. However, MolBarOpt still achieves superior convergence quality, with a final gradient norm of 6.50×10^{-8} Å^{-1}, several orders of magnitude lower than Newton-CG (3.64×10^{-1} Å^{-1}) and ANCOpt (2.18×10^{-3} Å^{-1}). For production applications of MolBar, where convergence quality is paramount to yielding a consistent identifier, MolBarOpt remains the preferred choice despite the increased wall time for very large systems.

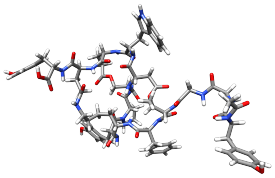
Based on these detailed evaluations, several clear recommendations apply to the implementation of structure unification in MolBar. For standard fragment sizes typical of drug-like molecules, MolBarOpt offers the optimal balance of computational speed, robustness, and convergence quality. Its integration of a backtracking line search guarantees reliable progression across highly diverse, steep-walled molecular topologies without necessitating the manual tuning of trust radii or step sizes. Conversely, the ANCOpt implementation—hampered by the RFO step control mechanism—exhibits recurrent convergence failures within the unification potential and is therefore not recommended for this application. For exceptionally large rigid systems (>500 atoms) characterized by highly constrained, rigid geometries, the choice between MolBarOpt and Newton-CG involves a trade-off. If absolute wall time is the overriding constraint and moderate residual gradients are deemed acceptable, Newton-CG may serve as a faster alternative due to the omission of the expensive ANC basis construction. However, because generating consistent, order-invariant molecular identifiers demands the highest possible structural convergence, MolBarOpt remains the preferred algorithm despite its increased computational overhead. Ultimately, the MolBarOpt algorithm successfully circumvents the unique optimization hurdles posed by the unification potential by harmonizing approximate normal coordinates with exact Hessian evaluations and flexible backtracking. Its exceptionally robust convergence logic guarantees that identical fragment graphs consistently relax into mathematically indistinguishable geometries, yielding perfectly matched eigenvalue spectra in the final topography matrices regardless of their starting conformers. This strict conformational invariance directly validates the underlying methodology of the unification procedure: it securely anchors the geometric relationships between substituents using standardized angle and dihedral constraints. Without possessing such a rigorous deterministic fallback, standard geometry optimizations would yield slightly varying topographies for identical substructures. Therefore, these benchmarking results empirically confirm that the MolBar fragmentation and unification protocol successfully extracts a truly conformer-independent representation of molecular fragments, while strictly preserving their 3D stereochemical identity. This enables the description of the relative configuration of each fragment in a consistent, unambiguous manner that is independent of the initial geometry and conformational state of the molecule.



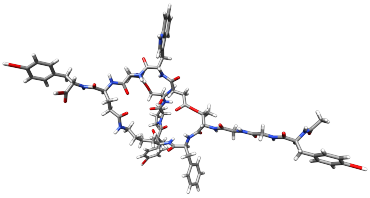
(a) Energy Convergence



(b) Gradient Norm Convergence



(c) Starting Structure

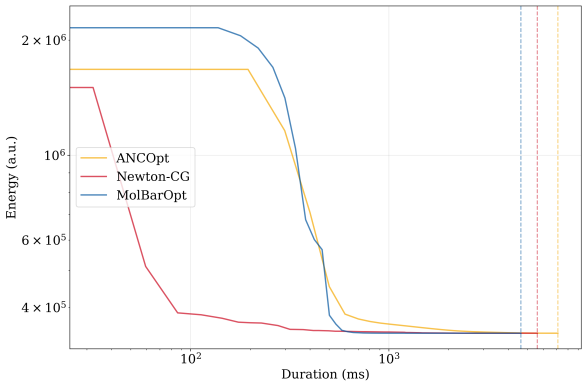


(d) Final Structure

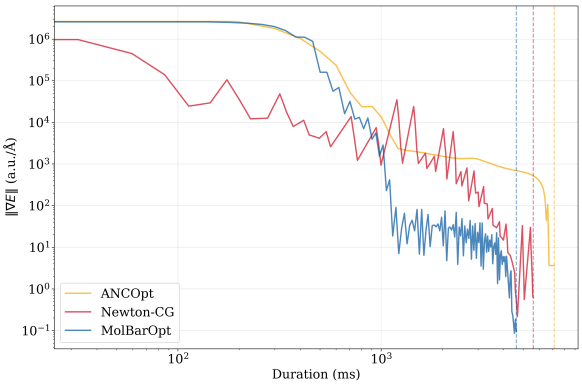
Figure 5.18: Performance comparison of three optimization algorithms for structure unification of Microviridin. The top row displays energy (a) and gradient norm (b) convergence as a function of wall time, with vertical dashed lines indicating completion times. The MolBarOpt algorithm employing line search with backtracking (blue) converges in 6.57 seconds, followed by the ANCOpt implementation (orange) at 7.22 seconds, while the Newton-CG method (red) requires 38.77 seconds. ANCOpt fails to fully converge with a final gradient norm of 4.07×10^3 Å^{-1} compared to MolBarOpt's 4.38×10^{-4} Å^{-1}. Newton-CG reaches a low final gradient norm of 5.58 Å^{-1} but takes significantly longer. The bottom row shows the initial molecular geometry (c) and the final idealized structure (d) obtained after optimization. The idealization process mostly corresponds to a stretching of the molecular structure, with ideal angles corresponding to VSEPR geometries and planarization of the ring systems.

Table 5.10: Optimizer Comparison Results for Microviridin

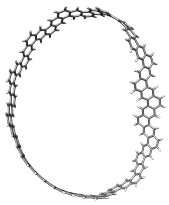
Metric	MolBarOpt	ANCOpt	Newton-CG
Wall time (s)	6.57	7.22	38.77
Iterations	364	173	795
Final energy (a.u.)	2.81×10^5	2.82×10^5	2.81×10^5
Final $\|\nabla\|$(Å^{-1})	4.38×10^{-4}	4.07×10^3	5.58
Final step norm (Å)	3.91×10^{-11}	3.50×10^{-1}	1.69×10^{-5}



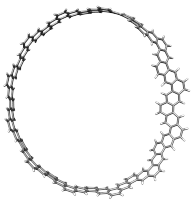
(a) Energy Convergence



(b) Gradient Norm Convergence



(c) Starting Structure

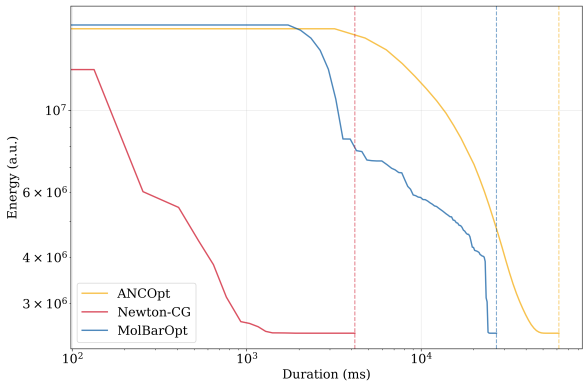


(d) Final Structure

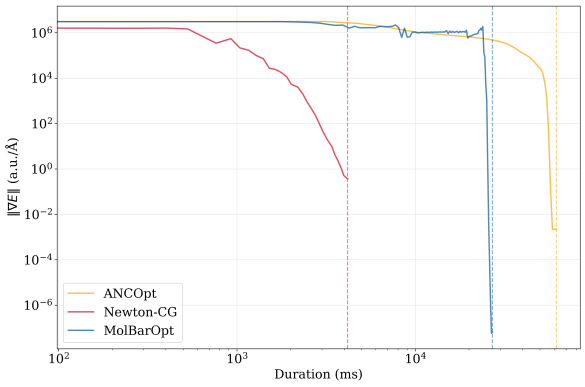
Figure 5.19: Performance comparison of three optimization algorithms for structure unification of a Möbius 50-cyclacene structure. The top row displays energy (a) and gradient norm (b) convergence as a function of wall time, with vertical dashed lines indicating completion times. The MolBarOpt algorithm employing line search with backtracking (blue) converges in 4.66 seconds, followed by the Newton-CG method (red) at 5.59 seconds, while the Python ANCOpt implementation (orange) requires 7.13 seconds. The line search approach achieves both faster convergence and a lower final gradient norm (9.39×10^{-2} Å^{-1}) compared to Newton-CG (6.19×10^{-1} Å^{-1}) and the RFO-based ANCOpt (3.64 Å^{-1}). The bottom row shows the initial molecular geometry (c) and the final idealized structure (d) obtained after optimization.

Table 5.11: Optimizer Comparison Results for Möbius 50-cyclacene

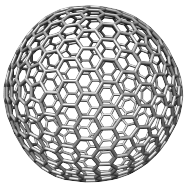
Metric	MolBarOpt	ANCOpt	Newton-CG
Wall time (s)	4.66	7.13	5.62
Iterations	110	70	84
Final energy (a.u.)	3.43×10^{5}	3.43×10^{5}	3.43×10^{5}
Final $\|\nabla\|$(Å^{-1})	9.39×10^{-2}	3.64	6.19×10^{-1}
Final step norm (Å)	4.30×10^{-8}	5.68×10^{-1}	4.34×10^{-6}



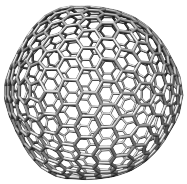
(a) Energy Convergence



(b) Gradient Norm Convergence



(c) Starting Structure



(d) Final Structure

Figure 5.20: Performance comparison of three optimization algorithms for structure unification of C_{720} fullerene. The top row displays energy (a) and gradient norm (b) convergence as a function of wall time, with vertical dashed lines indicating completion times. Interestingly, for this large system with 720 atoms, the Newton-CG method (red) achieves the fastest convergence at 4.21 seconds, followed by the MolBarOpt algorithm with line search (blue) at 27.15 seconds and the Python ANCOpt implementation (orange) at 61.71 seconds. This reversal in performance compared to smaller systems suggests that the overhead of coordinate transformations in ANC-based methods becomes more significant for very large systems. Despite the longer wall time, the MolBar implementation achieves the lowest final gradient norm (6.50×10^{-8} Å^{-1}) compared to ANCOpt (2.18×10^{-3} Å^{-1}) and Newton-CG (3.64×10^{-1} Å^{-1}), demonstrating superior convergence quality. The bottom row shows the initial molecular geometry (c) and the final idealized structure (d) obtained after optimization.

Table 5.12: Optimizer Comparison Results for C_{720}

Metric	MolBarOpt	ANCOpt	Newton-CG
Wall time (s)	27.15	61.71	4.21
Iterations	81	41	31
Final energy (a.u.)	2.49×10^6	2.49×10^6	2.49×10^6
Final $\|\nabla\|$(Å^{-1})	6.50×10^{-8}	2.18×10^{-3}	3.64×10^{-1}
Final step norm (Å)	1.80×10^{-13}	1.43×10^{-7}	5.69×10^{-5}

5.6 Absolute Configuration Matrix: Chirality Index and Priorities

With the molecule now represented as a collection of unified fragments, the focus shifts to characterizing the chirality of these fragments and their interactions. Therefore, the next step in the MolBar identifier generation process is to compute the absolute configuration matrix, which encapsulates the chirality information of the molecule at the fragment level. To assemble this matrix, the process begins with the construction of a fragment graph. Figure 5.21 provides an illustration of the fragment graph for 3-(3-*tert*-butylcyclobutylidene)piperidin-2-one.

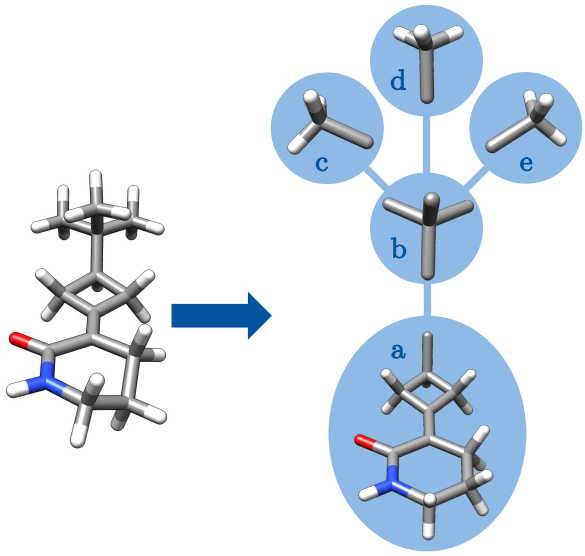


Figure 5.21: Fragment graph for 3-(3-*tert*-butylcyclobutylidene)piperidin-2-one. Each vertex $v_a \in V_{\text{fragments}}$ (blue circle) denotes a unified fragment. An edge $e_{ab} \in E_{\text{fragments}}$ indicates adjacency between fragments, reflecting an overlap in their atomic composition $v_i \in V_{\text{atoms}}$. Reprinted from van Staalduinen, N., Bannwarth, C., *MolBar: a molecular identifier for inorganic and organic molecules with full support of stereoisomerism*, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.

Here, each vertex $v_a \in V_{\text{fragments}}$ in this graph represents a fragment. An edge $e_{ab} \in E_{\text{fragments}}$ exists between two vertices whenever their corresponding fragments are adjacent, indicating the presence of overlapping atoms $v_i \in V_{\text{atoms}}$ between the fragments. Note that fragments are defined as overlapping. Fragments can be isolated; however, in such cases, they are not connected to any other fragment in the graph. For the chirality matrix, the chirality index G_a (Section 5.6.1) of each fragment v_a is determined, along with the priority p_a of each fragment (Section 5.6.2). Further, the distance d_{ab} between each pair of fragments v_a and v_b is calculated. This distance corresponds to the number of edges in the shortest path connecting the fragments and is calculated using the BFS algorithm (Section 2.3).

Due to the dependence of the chirality matrix on the results of earlier stages, particularly the unification process (Section 5.5), a thorough benchmarking of the chirality matrix is presented in Sections 6.2 and 6.5. There, the ability of the chirality matrix to distinguish be-

tween stereoisomers and contribute to the deduplication of a large dataset of organic molecules (which includes many chiral compounds) is evaluated.

5.6.1 Determining Absolute Configuration of Arbitrary Fragments

Every fragment v_a requires characterization through an absolute configuration index, denoted as G_a. This index must satisfy two fundamental conditions: it should equal zero for non-chiral fragments and take a non-zero value for chiral ones. Additionally, it must reverse its sign upon spatial inversion of a chiral fragment, thereby enabling the determination of absolute configuration.[249] Numerous chirality indices have been proposed over time. These include the Hausdorff measure,[250,251] the helicity tensor introduced by Ferrarini and Nordio,[252] and a measure proposed by Luzanov and Babich, which is derived from the mean torsion within a molecule.[253]

An alternative approach involves deriving a chirality index from optical activity theory. Since chiral molecules exhibit distinct specific rotation angles, it is reasonable to begin with an optical activity tensor to establish a geometric chirality measure. Osipov, Pickup, and Dunmur demonstrated that the pseudoscalar behavior of a chiral molecule, defined by its Cartesian coordinates $\mathbf{R}$, can be captured by the trace of the gyration tensor $\boldsymbol{\mathcal{G}}$ (Equation 5.62).[254] This trace is expressed as:

$$G_0(\mathbf{R}) = \frac{1}{3}\mathrm{Tr}\,\boldsymbol{\mathcal{G}}(\mathbf{R}) \tag{5.62}$$

$$\begin{aligned} G_0(\mathbf{R}) = &\int \rho(\mathbf{r}_1)\rho(\mathbf{r}_2)\rho(\mathbf{r}_3)\rho(\mathbf{r}_4) \\ &\times \frac{\big[(\mathbf{r}_{12}\times\mathbf{r}_{34})\cdot\mathbf{r}_{14}\big](\mathbf{r}_{12}\cdot\mathbf{r}_{23})(\mathbf{r}_{23}\cdot\mathbf{r}_{34})}{(r_{12}r_{23}r_{34})^{n_1}r_{14}^{n_2}} \\ &\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2\mathrm{d}\mathbf{r}_3\mathrm{d}\mathbf{r}_4, \end{aligned} \tag{5.63}$$

where $\mathbf{r}_{ij}$ denotes the distance vector between two atoms, and r_{ij} is its magnitude. $\rho(\mathbf{r}_i)$ describes an arbitrary density of the molecule of interest, while n_1 and n_2 are integer exponents that can be adjusted to modify the weight of distance contributions. When $n_1 = 2$ and $n_2 = 1$, the index becomes dimensionless, which is the setting adopted in this work. The isotropic chirality index G_0 remains invariant under rotation and translation. As a pseudoscalar, G_0 changes sign under spatial inversion for chiral objects, while it is zero exclusively for achiral molecules. This property allows G_0 to theoretically describe the chirality of a molecule, irrespective of whether it exhibits central, axial, or planar chirality. By replacing $\rho(\mathbf{r})$ with point masses p_i through delta functions $\delta(|\mathbf{r}-\mathbf{r}_i|)$, G_0 simplifies to its discrete form in Equation 5.64:

$$G_0(\mathbf{R}) = \sum_{ijkl} p_i p_j p_k p_l \frac{(\mathbf{r}_{ij}\mathbf{r}_{kl}\mathbf{r}_{il})(\mathbf{r}_{ij}\mathbf{r}_{jk})(\mathbf{r}_{jk}\mathbf{r}_{kl})}{(r_{ij}r_{jk}r_{kl})^{n_1}r_{il}^{n_2}}. \tag{5.64}$$

For the purpose of the identifier, only the sign of G_0 is relevant. Thus, the chirality index

G_a of a fragment, described by the Cartesian coordinates $\mathbf{R}_a^{\text{uni}}$ after the unification process, is defined as:

$$G_a = \begin{cases} 1 & \text{if } G_0(\mathbf{R}_a^{\text{uni}}) > 0, \\ 0 & \text{if } G_0(\mathbf{R}_a^{\text{uni}}) = 0, \\ -1 & \text{if } G_0(\mathbf{R}_a^{\text{uni}}) < 0. \end{cases} \tag{5.65}$$

However, Millar, Weinberg, and Mislow have highlighted a theoretical limitation: pseudoscalar functions used as chirality measures can yield a zero index even for inherently chiral objects.[249] Their argument posits that for any enantiomer $\mathbf{R}$, there exists a geometric distortion path transforming it into its enantiomer $\tilde{\mathbf{R}}$, where all intermediate structures along this path are chiral. Since G_0 must change sign, there must exist chiral structures for which $G_0 = 0$. Further research is necessary to determine the frequency of such occurrences, although current tests confirm the utility of this approach in describing the absolute configuration of unified fragments (see Section 5.4.4). The unification process, which prevents any geometric distortion path between $\mathbf{R}$ and $\tilde{\mathbf{R}}$ by standardizing each input fragment structure into either $\mathbf{R}_a^{\text{uni}}$ or $\tilde{\mathbf{R}}_a^{\text{uni}}$, may potentially mitigate the occurrence of chiral zeros. As a further safeguard, a point group symmetry analysis is performed on each fragment to identify any point groups that are inherently achiral. The chirality index, G_a, is only calculated using Equation 5.65 if the fragment is classified as possibly chiral based on its point group symmetry. In all other cases, G_a is set to zero, ensuring that the chirality index is only computed for fragments that are theoretically capable of being chiral. Symmetry analysis is conducted using the Python library `pymatgen`.[255]

5.6.2 Assignment of Atomic and Fragment Priorities

The methodology necessitates the introduction of pseudo masses p_i for each atom v_i as referenced in Equation 5.64. Relying solely on atomic masses m_i would incorrectly categorize identical atoms of the same element in distinct chemical environments as equivalent. Consider fragment b in Figure 5.21, where the four adjacent carbon atoms are not equivalent due to their differing substituents: three belong to a methyl group, while one is part of a ring system.

To address this, a pseudo mass p_i is defined, which is referred to as the atomic priority for each atom v_i. This priority is a numerical value that reflects the atom's identity and its chemical environment. Theoretically, this priority can adopt any form, provided it ensures that equivalent atoms are grouped together. In quantum chemistry, symmetry-equivalent atoms exhibit identical atomic populations and therefore identical partial charges within a given partitioning scheme. A common approach in quantum chemistry to determine the electron distribution among atoms in a molecule is the Mulliken population analysis.[17] This method partitions the total density into atomic contributions based on the coefficients of the atomic orbitals involved in each molecular orbital. The analysis is expressed as:

$$\begin{aligned}\sum_{k}^{n_{\text{orb}}} n_k \int \phi_k^2 \mathrm{d}\mathbf{r} &= \sum_{\mu\nu}^{n_{\text{basis}}} \Big(\sum_{k}^{n_{\text{orb}}} n_k c_{\mu k} c_{\nu k}\Big) S_{\mu\nu} \\ &= \sum_{\mu\nu}^{n_{\text{basis}}} P_{\mu\nu}^{\text{dens}} S_{\mu\nu} \\ &= \sum_{\mu}^{n_{\text{basis}}} (\mathbf{P}^{\text{dens}}\mathbf{S})_{\mu\mu} = \mathbf{Tr}(\mathbf{P}^{\text{dens}}\mathbf{S}) \\ &= n_{\text{elec}}\end{aligned} \tag{5.66}$$

Here, $c_{\mu k}$ represents the coefficients of the atomic orbitals, n_k the occupation number, $\mathbf{P}$ the density matrix, and $\mathbf{S}$ the overlap matrix. The term $(\mathbf{PS})_{\mu\mu}$ corresponds to the number of electrons associated with the basis function ϕ_μ. Summing these contributions over all basis functions for an atom v_i yields the density at that atom.

While Mulliken population analysis is typically applied in electronic structure theory, its computational cost renders it impractical for molecular identifiers like MolBar. Instead, an analogy to simple Hückel theory is made, where the eigenvectors of the Hückel matrix represent MOs. Similarly, the eigenvectors of the MolBar topography matrix serve as non-physical orbitals, encoding information about the molecule's 3D shape and symmetry. Diagonalizing the MolBar topography matrix (Equation 5.67) produces these MolBar orbitals and their associated eigenvalues:

$$\mathbf{H}_{\text{topography}}\mathbf{C} = \mathbf{EC} \tag{5.67}$$

These orbitals enable the calculation of an artificial density for each atom. However, determining the occupation of MolBar orbitals presents a challenge, as the topography matrix lacks physical meaning. In electronic structure theory, orbitals are populated according to the *Aufbau* principle, favoring the lowest eigenvalues for stability. In MolBar, orbitals with positive eigenvalues, which exhibit fewer nodal planes, are occupied. Each such orbital is assigned an arbitrary occupation number n_k of two, ensuring a consistent rule for degenerate orbitals. This process yields the MolBar density matrix, denoted as $\mathbf{P}^{\text{MolBar}}$:

$$\mathbf{P}^{\text{MolBar}} = \mathbf{CN}_{\text{occ}}\mathbf{C}^{\intercal} \tag{5.68}$$

Here, $\mathbf{N}_{\text{occ}}$ is the occupation matrix, indicating whether an orbital is occupied. Assuming zero overlap between orbitals, as in Hückel theory, the overlap matrix simplifies to the identity matrix. Consequently, the artificial atomic density for each atom is directly given by the diagonal elements of the density matrix:

$$P_{\mu\mu}^{\text{MolBar}} = \rho_i \tag{5.69}$$

Each atom $\{v_i\}$ is thus assigned an artificial density ρ_i. These densities are ranked to produce a set of atomic priorities $\{p_i\}$. The use of the topography matrix ensures that atomic priorities reflect connectivity and even subtle differences, such as variations in double bond configurations. Equivalent atoms share the same density and priority, while non-equivalent atoms receive distinct priorities.

For the absolute configuration matrix (see Section 4.3), fragment priorities $\{p_a\}$ are required rather than atomic priorities. Fragment priorities are determined by comparing the highest atomic priorities within each fragment (see Figure 5.22). The fragment with the highest atomic priority is assigned the highest fragment priority. In cases where multiple fragments tie for the highest priority, the comparison proceeds to the second-highest priority, continuing until all fragments are differentiated or deemed equivalent.

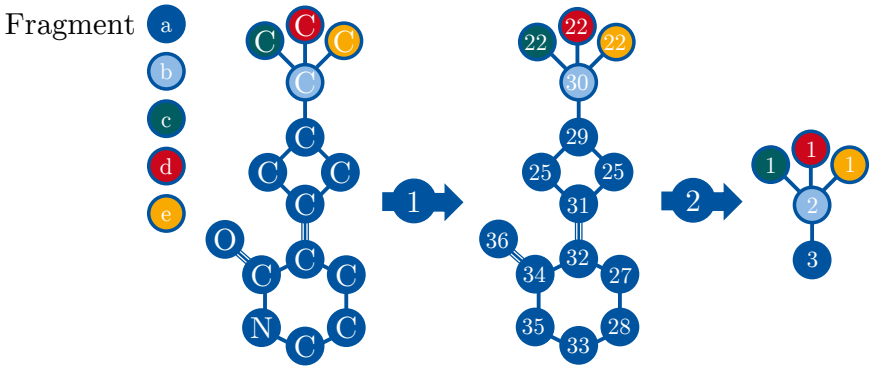


Figure 5.22: (1) Each fragment is assigned a priority based on the atomic priorities of its constituent atoms. (2) The highest atomic priority within a fragment is compared to that of another fragment, with the fragment containing the highest atomic priority receiving the highest fragment priority. Reprinted from van Staalduinen, N., Bannwarth, C., *MolBar: a molecular identifier for inorganic and organic molecules with full support of stereoisomerism*, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.

5.7 Structure of the Molecular Barcode (MolBar) Identifier

MolBar employs a collection of the aforementioned matrices: topology (Section 5.2), topography (Section 5.5), and chirality (Section 5.6). Specifically, the eigenvalue spectra of these matrices are utilized to create a unique identifier for each molecule. This approach allows MolBar to encode the topology, topography, and absolute configuration of a molecule in a compact and efficient manner. The representation of the latter two matrices depends on a partitioning process that divides the molecule into rigid substructures, enabling a consistent 3D depiction within the conformational space. Earlier sections outlined algorithms for defining the molecular graph $E(G)$, the set of electronically hindered bonds $M_{\text{electronic}}$, the set of sterically hindered bonds M_{steric}, the idealized distances R_{ij} following structure unification, the chirality index G_a for each rigid fragment, and the priorities p_a required to construct MolBar. With these properties established, the various matrices can be assembled, including the topology matrix $\mathbf{H}^{\text{topology}}$ (Equation 4.1), the topography matrix $\mathbf{B}^{\text{topography}}$ (Equation 4.4), and the chirality matrix $\mathbf{G}^{\text{chirality}}$ (Equation 4.8). The overall MolBar generation process, from the input of the 3D structure to the final identifier, is illustrated in a flowchart for 3-(3-*tert*-butylcyclobutylidene)piperidin-2-one (Figure 5.23).

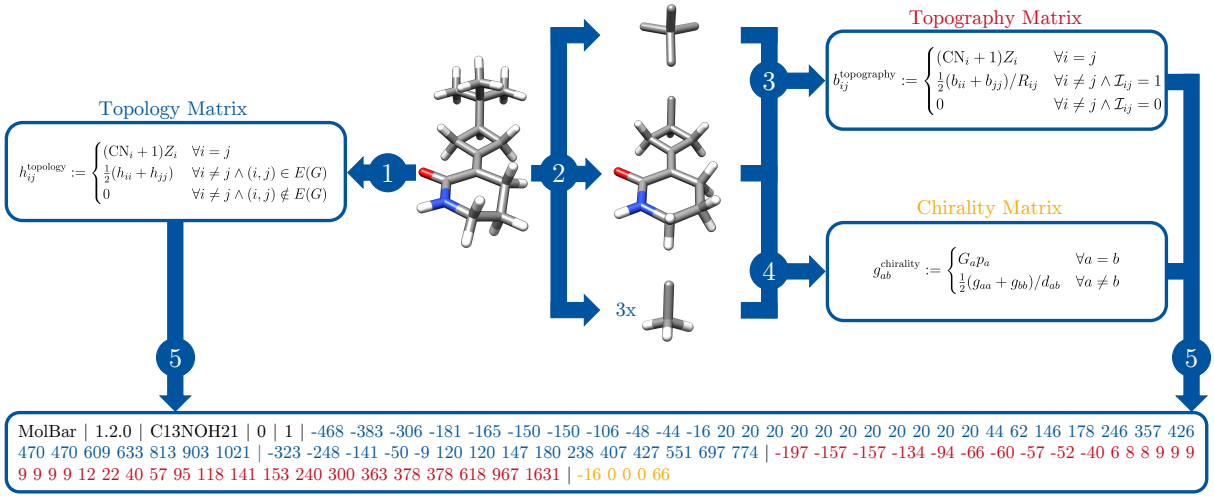


Figure 5.23: Flowchart illustrating the MolBar generation process. The molecular topology is represented using an extended adjacency matrix, comparing Cartesian distances with reference values (1). The molecule is then partitioned into rigid fragments, which are unified using a specialized force field (2). A modified Coulomb matrix describes the 3D atomic arrangement (3), while Osipov-Pickup-Dunmur indices determine the absolute configuration (4). Finally, the eigenvalue spectra of all matrices are scaled, rounded, and concatenated along with the identifier name, version, chemical formula, charge, and multiplicity (5). Modified from van Staalduinen, N., Bannwarth, C., *MolBar: A molecular identifier for inorganic and organic molecules with full support of stereoisomerism*, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission of the Royal Society of Chemistry under the CC-BY 3.0 license.

Initially, the molecular topology is represented by an extended adjacency matrix (Equation 4.1), which incorporates details about bond presence and types. The molecule is then partitioned into rigid fragments based on bond orders and ring structures, followed by a structure uni-

fication step to ensure a consistent 3D representation of these fragments. Subsequently, the topography matrix (Equation 4.4) is constructed to capture the 3D arrangement of atoms in the unified fragments, while the chirality matrix (Equation 4.8) encodes the absolute configuration of these fragments using their Osipov-Pickup-Dunmur indices. The final MolBar identifier is generated by diagonalizing these matrices and concatenating their eigenvalue spectra to produce a unique molecular identifier. The eigenvalues are scaled by a factor of ten and rounded to the nearest integer to minimize the size of the MolBar, as prior results have demonstrated that additional precision is unnecessary. The complete MolBar for 3-(3-*tert*-butylcyclobutylidene)piperidin-2-one is provided in Table 5.13.

Table 5.13: Structure of MolBar as in the example for chiral 3-(3-*tert*-butylcyclobutylidene)piperidin-2-one.

MolBar	Example
version	1.2.0
molecular formula	C13NOH21
total molecular charge	0
multiplicity	1
topology	-468 -383 -306 -181 -165 -150 -150
	-106 -48 -44 -16 20 20 20 20 20 20 20 20
	20 20 20 44 62 146 178 246
	357 426 470 470 609 633 813 903 1021
heavy atom topology	-323 -248 -141 -50 -9 120 120 147
	180 238 407 427 551 697 774
topography	-197 -157 -157 -134 -94 -66 -60 -57 -52 -40
	6 8 8 9 9 9 9 9 9 9 12 22 40 57 95
	118 141 153 240 300 363 378 378 618 967 1631
chirality	-16 0 0 0 66

The initial segment includes the version number, molecular formula, total molecular charge, and multiplicity. Specifying the version ensures that only MolBar identifiers of the same version are compared, as future updates or corrections may alter the resulting spectra. If custom bond orders, adjusted thresholds for hindered bond detection, or specific bond hindrance details are provided by the user, the version number is appended with an 'M' to indicate modification. This adjustment is essential, since manual modifications disrupt the one-to-one correspondence between a 3D structure and its MolBar identifier, which is a core principle of the MolBar framework. The eigenvalue spectra are concatenated in a defined sequence:

$$\lambda_{\text{topology}}, \lambda_{\text{topology},Z_i>Z_{\text{H}}}, \lambda_{\text{topography}}, \lambda_{\text{chirality}}.$$

First, the eigenvalue spectrum $\lambda_{\text{topology}}$ of the topology matrix (Equation 4.1) is presented,

followed by the spectrum of the topology matrix excluding hydrogen atoms. The topology matrix excluding hydrogen atoms uses the coordination numbers on the diagonal minus those of the adjacent hydrogen atoms. Including hydrogen atoms in one matrix and omitting them in another enables the differentiation of prototropic tautomers, as these molecules share the same heavy atom topology spectrum but differ in their total topology spectrum. By default, tautomeric structures are assigned distinct MolBar identifiers, which is beneficial for quantum chemical structure databases, as it ensures each structure is uniquely represented based on its electronic properties. While this approach to tautomerism is simplistic and does not account for significant energy barriers in hydrogen shifts,[256] it remains practical for the purposes of quantum chemical structure databases. The topography spectrum $\lambda_{\text{topography}}$ (Equation 4.4) is then provided, followed by the absolute configuration spectrum (Equation 4.8).

The `molbar` implementation is accessible as an open-source Python module, with resource-intensive operations handled by Fortran subroutines. The program package, along with its documentation, is available at `https://git.rwth-aachen.de/bannwarthlab/molbar`. Installation is possible via the Python Package Index (PyPI) using the command `pip install molbar`. The package accepts 3D molecular coordinates as input and supports a range of file formats detailed in the documentation.

6 Results and Discussion: Benchmarking MolBar and Application to Chemical Problems

Sections 6.2 to 6.6 of this chapter are based on the following publications:

- van Staalduinen, N., Bannwarth, C., MolBar: a molecular identifier for inorganic and organic molecules with full support of stereoisomerism, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.
- van Staalduinen, N., Bannwarth, C., When Bonds Resist Rotation: Automated Classification and Representation of Atropisomers in the MolBar Identifier, *J. Chem. Inf. Model.*, **2026**, *66*, 10935–10955, DOI: `10.1021/acs.jcim.6c00688`, with permission from the American Chemical Society.

The author of this thesis is also the primary contributor to the original publications, with the only other author being the supervising professor as corresponding author. Many sections have been significantly revised, as all content has been updated to reflect the latest version of the MolBar software (1.2.0), which has undergone continuous development and improvement since the original publications. Additionally, different datasets and benchmark molecules were employed in numerous instances, and the results have been updated to align with these changes. The most significant modifications to the original publications are outlined below:

- Section 6.2: The same example molecules from the original publication were retained, but the results were updated to correspond with MolBar 1.2.0. Furthermore, these results were integrated with the example molecules concerning atropisomerism from the second publication from 2026.
- Section 6.3: Unlike the original 2024 publication, which included a permutation invariance test, this thesis uses a different dataset and visualizes the test by comparing MolBar's permutation invariance with a permutation-variant property.
- Section 6.4: While the original 2024 publication benchmarked the computational cost of the MolBar identifier, this thesis employs a different dataset and compares the cost of computing MolBar with that of the InChI identifier and single-point calculations at the GFN2-xTB and UMA levels of theory. The computational cost of MolBar is also analyzed in greater detail by breaking down the cost into distinct algorithmic steps, an analysis absent in the original publication.

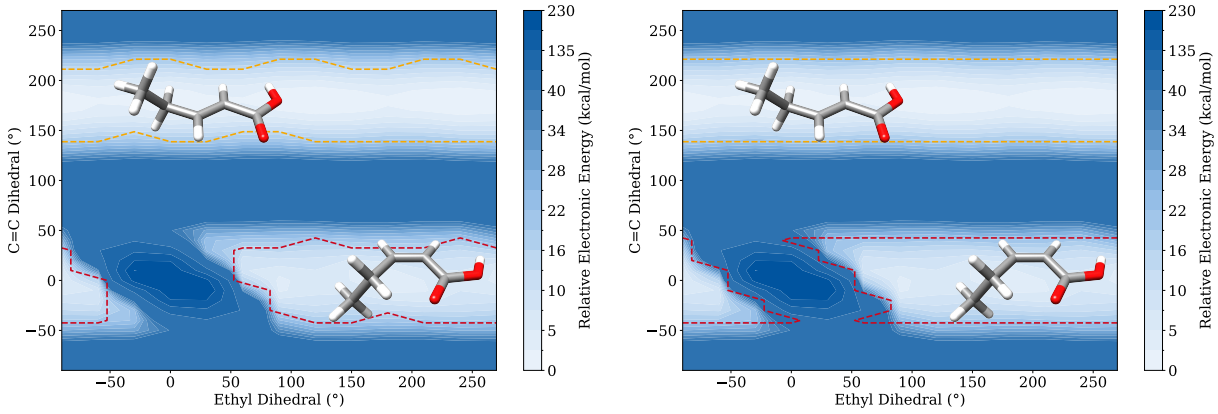
- Section 6.5: The original 2024 publication presented deduplication results for a large dataset of organic molecules, whereas this thesis uses a different dataset with more high-quality structures and updates the results to reflect MolBar 1.2.0.
- Section 6.6: The original 2024 publication included a molecular dynamics (MD) simulation, but this thesis uses a different molecule and updates the results to reflect MolBar 1.2.0. Analyses of conformer ensembles and reaction networks, not present in the original publication, have also been added.

6.1 Mapping the Potential Energy Surface through MolBar

The goal of this work was to develop a novel molecular identifier, MolBar. This identifier is intended to map and classify stationary points on the PES, distinguishing regions on its surface that are separated by sufficiently high energetic barriers. MolBar is tailored specifically for quantum chemical applications. It determines molecular identity directly from 3D coordinates, removing reliance on external tools and their subjective bond interpretations. This method ensures uniqueness and consistency for any provided coordinate set. MolBar represents the molecule using three separate matrices. The topology matrix defines atom connectivity, allowing the description of constitutional isomerism. The topography matrix, however, encodes the spatial arrangement of atoms in 3D, enabling differentiation between stereoisomeric forms. This matrix uses inverse distances between atoms, since these distances differ among diastereomers. It only includes distances within rigid fragments, which are characterized by limited conformational flexibility. This limitation prevents conformational changes from affecting the molecular description, as distances that vary significantly between conformers mainly convey conformational rather than identity-related information. Rigidity in substructures is assessed through analysis of ring structures, electronic hindrance (e.g., WBO analysis), and steric hindrance. As a general rule, MolBar employs a 20 kcal/mol threshold for dihedral changes to determine whether a substructure is rigid. To distinguish between left- and right-handed coordinate systems, the chirality matrix includes chirality indices for each fragment. This enables the description of central, axial, helical, and planar chirality. The final MolBar identifier is formed by concatenating the eigenvalue spectra of the respective matrices.

In the motivation for this work (see Chapter 3), approximated PES mappings for 2-pentenoic acid, isobutyramide, a dibromo-biphenyl-thioamide derivative, and a copper(I) complex with two phenanthroline ligands were discussed. The desired PES mapping is based on 20 kcal/mol basins, determined by calculating the energy required to reach an arbitrary point on the PES from each minimum along the MERP. These mappings were presented alongside the InChI-based PES mapping. The following section discusses the MolBar mappings achieved and compares them with the desired PES mappings. Each example uses an approximate PES calculated on a grid of two internal coordinates, with MLIP UMA-S[65] used for electronic energy computations. Since coordinates remain fixed and no relaxation occurs, the resulting reaction energies and activation barriers are illustrative rather than precise. MolBar was computed using the `molbar` Python package (version 1.2.0) with the function `get_molbar_from_coordinates`. It is important to note that the 20 kcal/mol basins serve only as a general guideline. Connectivity was not adjusted based on dissociation energies and the 20 kcal/mol threshold. While dihedral changes due to electronic and steric effects were fitted to produce different MolBar identifiers when the barrier between two minima exceeds 20 kcal/mol, MolBar does not determine the specific dihedral angle at which the 20 kcal/mol threshold is reached. Thus, the basins should be compared qualitatively rather than quantitatively. The key requirement is that two minima are assigned different MolBar identifiers when the barrier exceeds this value.

First, 2-pentenoic acid is presented in Figure 6.1, illustrating E/Z isomerism across the double bond. The PES is shown as a function of the double bond C-C=C-C dihedral and the ethyl group C-C-C=C dihedral. The target PES mapping in Figure 6.1a distinguishes between these isomers, while conformers related to ethyl group orientation are grouped. The MolBar PES mapping in Figure 6.1b differentiates between the E and Z isomers and aggregates conformers with varying ethyl orientations. Rigidity analysis divides 2-pentenoic acid into five rigid fragments: the CH_3 group, the C_2H_2 group, the double bond with its four substituents, the carbonyl group C-C(=O)O, and the hydroxyl group. As the double bond forms one rigid fragment, the topography matrix (Equation 4.4) captures the changing interatomic distances of the substituents for the E and Z forms. Since the CH_3 group and the C_2H_2 group are separated, no conformational changes are described.



(a) Target PES representation

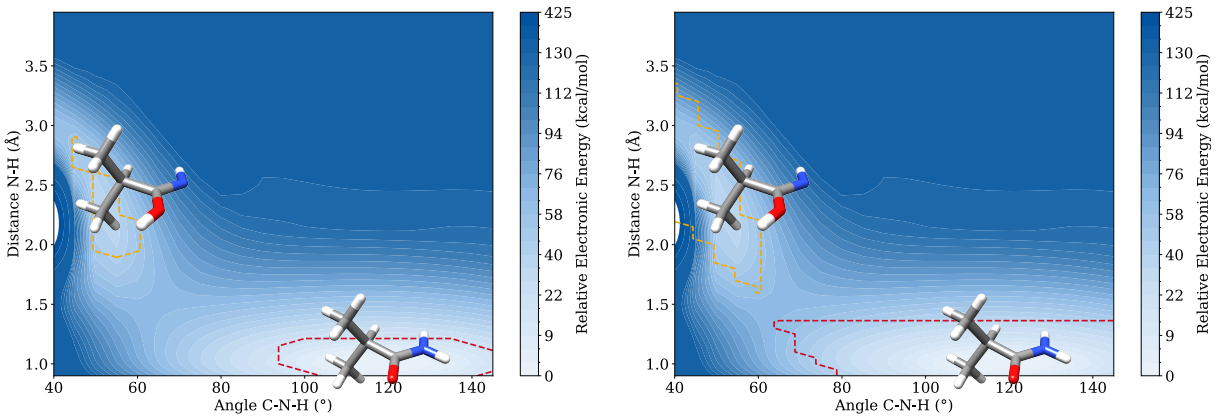
(b) MolBar-based PES representation

Figure 6.1: PES for the E/Z isomerization of 2-pentenoic acid. Energies are calculated using the MLIP UMA-S.[65] MolBar identifiers are generated with version 1.2.0 of the `molbar` Python package.[189] The E and Z isomers are distinct minima on the PES, separated by a significant barrier. The target mapping (a) distinguishes between isomers, aggregating ethyl-oriented conformers. The MolBar mapping (b) aligns with the target, differentiating E and Z isomers while grouping conformers with varying ethyl orientations.

Second, Figure 6.2 addresses the tautomerization of isobutyramide to its imidic acid form. The activation barrier for the hydrogen transfer from nitrogen to oxygen exceeds 20 kcal/mol, so the desired mapping separates both forms. While InChI merges these forms through tautomer canonicalization (since both forms coexist in solution on laboratory time scales), isobutyramide and its imidic acid remain distinguishable with distinct properties on quantum mechanical time scales. MolBar does not perform energy evaluation but assesses covalent radii for connectivity analysis. It recognizes the breaking of one covalent bond and the formation of another, resulting in a change in the topology matrix, thereby distinguishing between the tautomers.

Third, Sections 5.3 and 5.4 introduced electronic and steric hindrance classifiers in MolBar. These classifiers enable the identification of dihedral changes associated with a barrier of at least 20 kcal/mol, allowing MolBar to describe atropisomerism in (thio)-amides or biaryl sys-

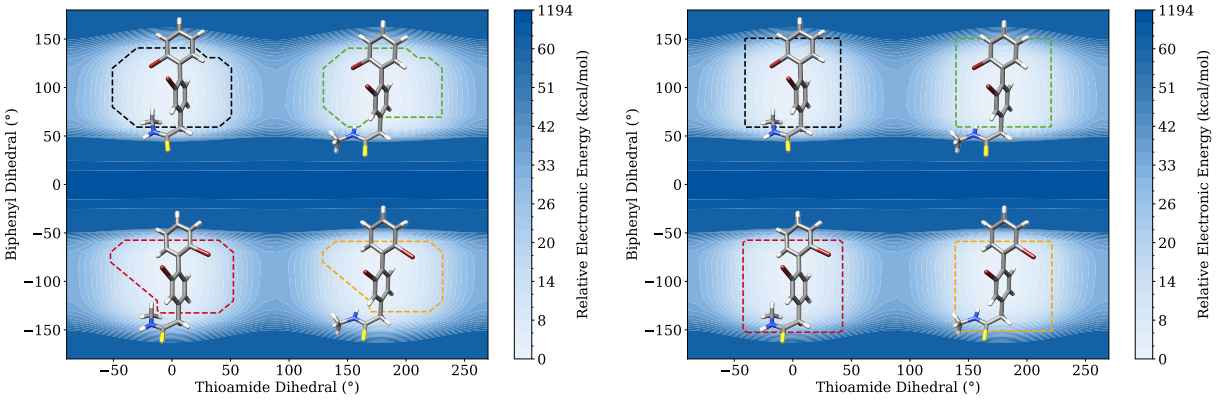
tems, among others. Figure 6.3 illustrates a dibromo-biphenyl-thioamide derivative, where all four atropisomers are distinct minima on the PES, separated by high energy barriers.



(a) Target PES representation **(b)** MolBar-based PES representation

Figure 6.2: PES for the tautomerization of isobutyramide to its imidic acid form. Energies are computed using the MLIP UMA-S.[65] MolBar identifiers are generated with version 1.2.0 of the `molbar` Python package.[189] The target mapping (a) separates the two tautomers. MolBar distinguishes between them due to changes in the topology matrix (Equation 4.1).

The biphenyl rotation barrier stems from steric hindrance between bromine atoms, while the thioamide C-N bond rotation barrier results from partial double-bond character due to resonance. The desired PES mapping (a) distinguishes between all four atropisomers. MolBar successfully differentiates between them (b), incorporating electronic hindrance (Section 5.3) and steric hindrance (Section 5.4) classifiers. The thioamide bond, with a WBO exceeding 1.25, is recognized by the electronic hindrance classifier, and its four substituents form one rigid fragment. Thus, the topography matrix varies for both thioamide configurations.

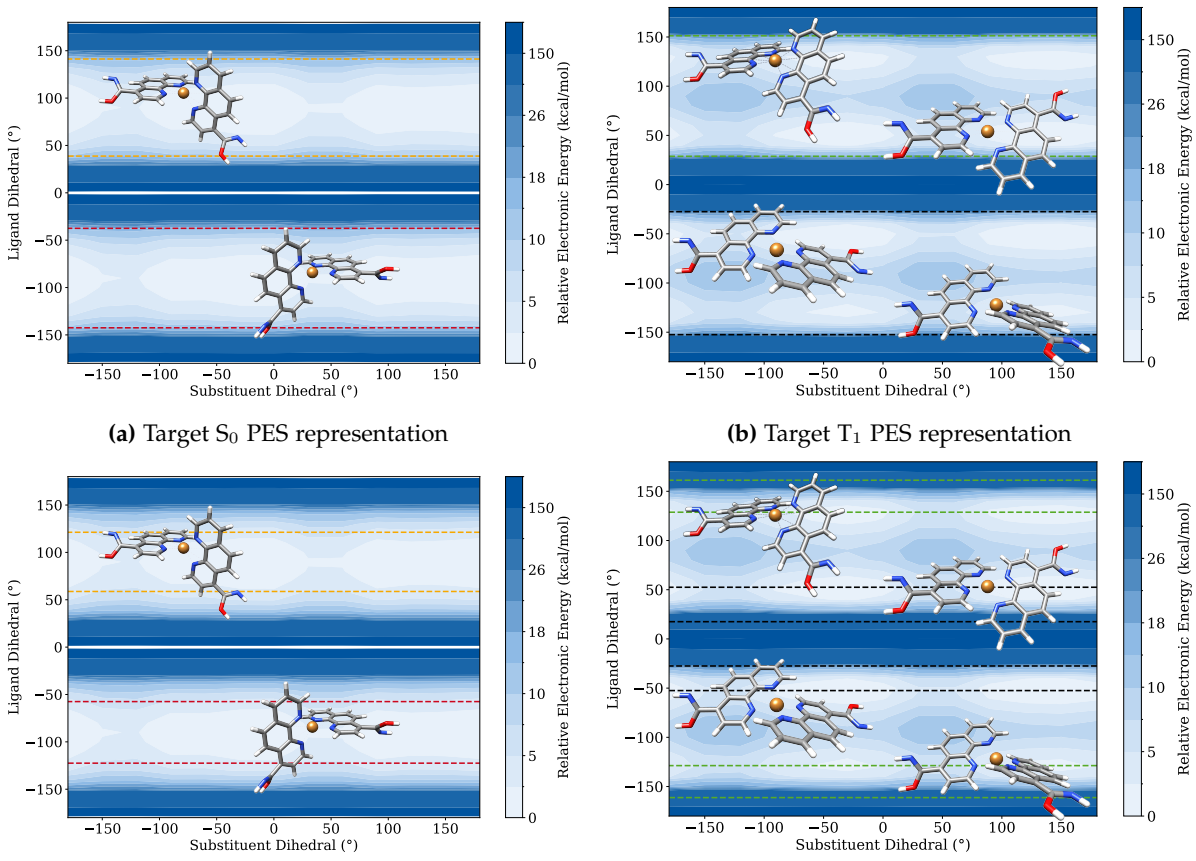


(a) Desired PES mapping **(b)** MolBar PES mapping

Figure 6.3: PES for atropisomerism in a dibromo-biphenyl-thioamide derivative. Energies are calculated using MLIP UMA-S.[65] MolBar identifiers are generated with version 1.2.0 of the `molbar` Python package.[189] All four atropisomers are distinct minima, separated by high barriers. MolBar differentiates them via electronic and steric hindrance classifiers, with changes in the topography and chirality matrices.

The steric hindrance classifier detects the high steric stress and high rotational barrier of the coplanar biaryl form. The biaryl substructure forms a single, chiral rigid fragment composed of two orthogonal rings. The absolute configuration of this fragment is described in the chirality matrix in Equation 4.8.

The final example shifts the focus from organic ground-state chemistry to inorganic excited-state systems. Figure 6.4 presents the PES for the ground (S_0) and first triplet excited (T_1) states of a copper(I) complex with two phenanthroline ligands. In the S_0 state, the PES minimum corresponds to a pseudo-tetrahedral geometry, where MolBar distinguishes between the two enantiomers using the chirality index and chirality matrix (Equation 4.8). For the T_1 state, the PES minimum adopts a near-square planar geometry, which can exist in either *cis* or *trans* arrangements. If one considers the local minima without evaluating the energetic barriers of the internal dihedral changes, the T_1 PES would exhibit four distinct minima: the *cis* and *trans* arrangements, each existing as a pair of enantiomers.



(a) Target S_0 PES representation

(b) Target T_1 PES representation

(c) MolBar-based S_0 PES representation

(d) MolBar-based T_1 PES representation

Figure 6.4: PES for ground (S_0) and first triplet excited (T_1) states of a copper(I) complex with two phenanthroline ligands. Energies are calculated using MLIP UMA-S.[65] MolBar identifiers are generated with version 1.2.0 of the `molbar` Python package.[189] S_0 features a distorted tetrahedral minimum with distinguishable enantiomers. T_1 exhibits a square planar minimum with *cis* and *trans* arrangements, where MolBar differentiates between local geometries despite low-energy transitions.

In T_1, the tetrahedral configuration linking the near-planar *cis* and *trans* structures represents a low-barrier transition state, falling well below MolBar's 20 kcal/mol threshold. High energy barriers only occur when both ligands are coplanar. Consequently, the desired PES mapping (Figures 6.4a and 6.4b) should distinguish the two T_1 enantiomeric sides (positive vs. negative ligand dihedral), while grouping the *cis* and *trans* arrangements together.

This represents the only case where MolBar does not fully align with the target PES mapping. Although MolBar is designed to aggregate structures separated by low energetic barriers, it erroneously differentiates between the *cis* and *trans* arrangements in the T_1 state. This discrepancy arises because MolBar simply distinguishes between the distinct local geometries without formally evaluating the low barrier associated with their interconversion. Furthermore, the structure idealization step optimizes the T_1 minimum into a perfectly square planar geometry. This introduces artificial symmetry and erases the chirality information of the minima. As a result, MolBar ends up distinguishing only the *cis* and *trans* arrangements—which should actually be grouped together—while failing to differentiate the two T_1 enantiomeric sides, which should remain separated.

Overall, MolBar successfully maps the PES for these examples, distinguishing between minima separated by high energy barriers while grouping together conformers with low barriers. These results demonstrate MolBar's capability to capture the desired PES mappings for a range of chemical systems and phenomena, including E/Z isomerism, tautomerization, atropisomerism, and excited-state geometries. This makes MolBar useful for mapping and classifying stationary points on the PES in quantum chemical databases and workflows, providing a consistent and unique molecular identifier based on 3D coordinates. In general, MolBar fulfills the goal of mapping the PES based on 20 kcal/mol basins, with only one minor deviation in the copper complex example.

6.2 Representation of Stereoisomers beyond Atom-Centered Chirality

Previously, MolBar was employed to describe four instances of double bond isomerism, tautomerism, atropisomerism, and distinct configurations of a copper complex. This section demonstrates the application of MolBar across a broader spectrum of stereoisomers. Examples include planar and axial chirality, as well as stereoisomers involving inorganic molecules. MolBar diverges from the traditional atomic representations used in SMILES and InChI, which center on atom-based stereochemistry. Instead, it aims to define the relative and absolute configuration of entire molecular fragments. Thus, MolBar is theoretically capable of representing a diverse array of molecules, encompassing both organic and inorganic types, and extending beyond the constraints of centrochirality. The following discussion explores cases where the atomistic description encounters limitations, drawing some examples from the tmQM dataset.[57] For the metal complexes, the InChI identifiers were sourced from the webdemo for inorganic molecules provided by the InChI Trust.[87] The input for the webdemo consisted of a Molfile generated using `OpenBabel` version 3.1.1.[85] It is important to note that these InChI identifiers do not yet conform to the standard InChI format, as indicated by the "InChI=1B" prefix. Instead, they represent future versions designed to incorporate metal-organic compounds, though they currently lack stereochemistry information.

Figure 6.5 illustrates stereoisomers involving both inorganic and organic molecules, alongside instances of planar and axial chirality. The first example (Figure 6.5a) presents two stereoisomers of a tungsten complex with octahedral coordination geometry. In MolBar, the central fragment corresponds to the octahedron itself. The distinct ligand arrangements within the octahedron differentiate the stereoisomers, particularly through the exchange of a carbon monoxide ligand with a hydroxide ligand. Consequently, variations in interatomic distances within the topography matrix enable MolBar to distinguish between the two stereoisomers. A similar situation arises in Figure 6.5b, where a square-planar platinum complex exhibits *cis*/*trans* isomerism. The stereoisomers in Figure 6.5c depict an octahedral osmium complex featuring two bidentate dithiocarbamate ligands and two phosphine ligands. Depending on ligand arrangement, the complex can be either chiral or achiral. Helical chirality emerges when the two bidentate ligands are not coplanar. In such cases, the complex comprises a single chiral fragment—the octahedron with the metal atom at its center. The absolute configuration spectra of the two enantiomers differ solely in sign. If both bidentate ligands lie in the same plane, the complex is achiral. The achiral variant displays different relative ligand positions within the octahedron compared to both chiral complexes, as reflected in the topography spectrum. Achirality is represented by the absolute chirality spectrum, which consists entirely of zeros.

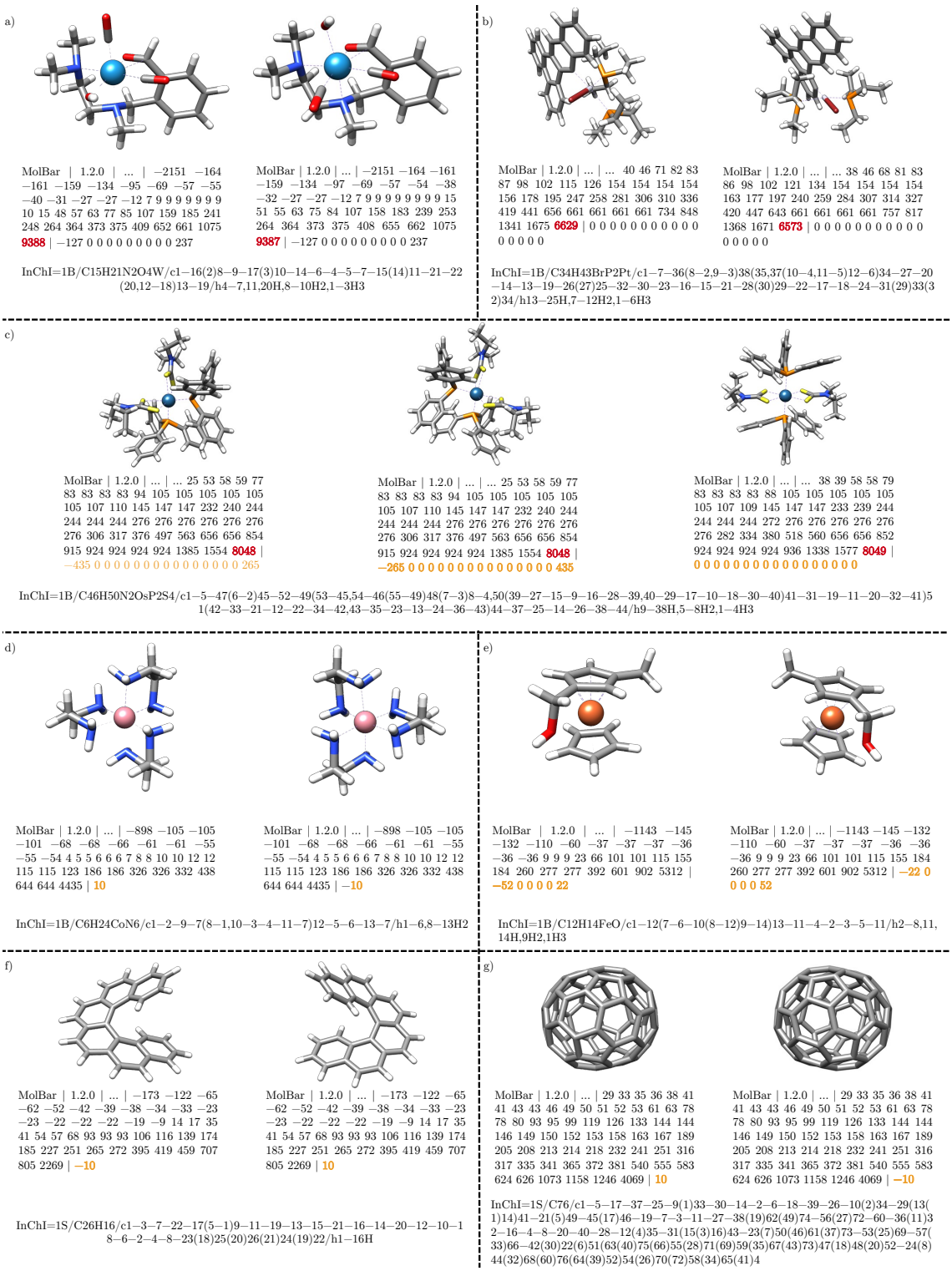


Figure 6.5: Examples demonstrating MolBar's capacity to distinguish stereoisomers beyond atom-centered stereochemistry: (a) two stereoisomers of a tungsten complex differing in ligand positions within the octahedron, (b) *cis*/*trans* isomerism in a square-planar platinum complex, (c) chiral and achiral configurations in an octahedral osmium complex, (d) Δ/Λ isomerism in tris(ethylenediamine)cobalt(III), (e) planar chirality in metallocenes, (f) axial chirality in hexahelicene, (g) axial chirality in fullerene C_{76}. Modified from van Staalduinen, N., Bannwarth, C., *MolBar: A molecular identifier for inorganic and organic molecules with full support of stereoisomerism*, *Digit. Discovery*, **2024**, *3*, 2298–2319, with permission from the Royal Society of Chemistry under the CC-BY 3.0 license.

Figure 6.5d showcases the classic example of Δ/Λ isomerism in tris(ethylenediamine)cobalt(III). Unlike the previous case, the entire molecule corresponds to a single fragment characterized by a unique eigenvalue in the absolute configuration spectrum. Another stereogenic unit can be observed in metallocenes, as illustrated in Figure 6.5e, where the iron atom coordinates to one side of a suitably substituted aromatic ring, resulting in planar chirality.

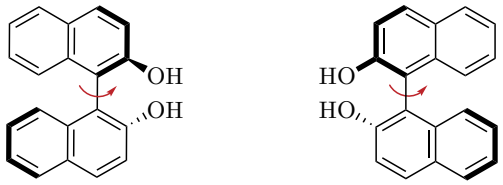
The chirality of hexahelicene in Figure 6.5f arises from the tendency to alleviate steric strain, leading to a right- or left-handed helical form within the polyaromatic ring system. A further notable example is the chiral fullerene C_{76} (Figure 6.5g), which exists in two enantiomeric forms. The absolute configuration of hexahelicenes and fullerenes is accurately described by the Osipov-Pickup-Dunmur index, which characterizes a single fragment and enables differentiation of the respective enantiomers using MolBar.

MolBar incorporates classifiers for both steric and electronic hindrance around single bonds.[154] This capability enables the representation of atropisomeric molecules. Figure 6.6 showcases a range of molecules with rotationally hindered single bonds, along with their MolBar representations. MolBar applies a threshold of 20 kcal/mol to classify rotationally hindered bonds. In the figure, atropisomeric axes with a predicted rotational barrier exceeding this threshold are highlighted in red, while those below the threshold are marked in blue. These predictions are based exclusively on the 3D coordinates of the molecule, including stereochemistry, without requiring quantum chemical calculations or manual input. The examples presented in this section were not part of the training, validation, or test sets discussed in Section 5.3 or 5.4. The first four examples demonstrate atropisomerism arising from steric hindrance. The final two examples serve to illustrate the electronic hindrance classifier. The InChI identifiers for these examples were obtained from the PubChem database for comparative purposes.

Figure 6.6a features 1,1'-bi-2-naphthyl, a classic atropisomer frequently used as a benchmark in stereochemical research and as a precursor to the BINAP/BINOL ligand family. The rotational barrier for this molecule has been measured at approximately 24 kcal/mol in benzene and dimethylformamide at 44°C and 50°C, respectively.[257] The steric hindrance classifier accurately predicts rotational hindrance, resulting in a large fragment that includes both naphthyl rings. The absolute configuration is determined for each enantiomer, as indicated by the orange eigenvalues in the MolBar representation. True enantiomers share identical topography eigenvalues but differ in their absolute configuration eigenvalues, specifically by a sign inversion (e.g., -4 0 14 becomes -14 0 4). A similar pattern is observed in Figure 6.6b, which depicts Murrastifoline F, an N,C-bonded heterobiarylic biscarbazole alkaloid isolated from *Murraya koenigii*. This molecule is classified as sterically hindered, with AM1 calculations indicating a rotational barrier of 39 kcal/mol.[258] Both enantiomers differ only in the absolute configuration eigenvalues associated with the atropisomeric axis. Figure 6.6c presents Sotorasib, a covalent $KRAS^{G12C}$ inhibitor approved by the FDA in 2021 for the treatment of non-small-cell lung cancer.[203] Experimental studies using time-course NMR report a rotational barrier exceeding 33.5 kcal/mol. The molecule contains three potential atropisomeric axes, but only

one is identified as sterically hindered by MolBar. Rotation around this axis produces a diastereomer rather than an enantiomer due to the presence of another stereocenter, leading to differences in the magnitude of the absolute configuration eigenvalues rather than their sign. Marinopyrrole A, shown in Figure 6.6d, is a marine *Streptomyces* metabolite with promising antibiotic activity against methicillin-resistant *Staphylococcus aureus* (MRSA). In the absence of additional stereoelements, the absolute configuration eigenvalues differ in sign. Figure 6.6e illustrates a thiolamide characterized by a strong partial double bond in its C-N bond. The computational protocol described in Section 5.4 yields a rotational barrier of 23.6 kcal/mol, consistent with literature reports of barriers exceeding 20 kcal/mol.[120] While steric hindrance is not present, the electronic hindrance classifier predicts a WBO above 1.25, with a calculated WBO of 1.3 at the GFN2-xTB level. Thus, this axis is classified as rotationally hindered. In contrast, Figure 6.6f shows paracetamol, a widely used analgesic and antipyretic drug that is not considered atropisomeric.[259] Calculations yield a rotational barrier of 19.5 kcal/mol, below the 20 kcal/mol threshold, and a C-N bond WBO of 1.19 at the GFN2-xTB level, which is below the threshold of 1.25. This aligns with the prediction of the electronic hindrance classifier. Consequently, MolBar does not classify paracetamol as rotationally hindered and generates a single representation regardless of the input rotamer. These examples underscore MolBar's proficiency in distinguishing atropisomeric from non-atropisomeric molecules and capturing both steric and electronic contributions to rotational barriers directly from molecular geometry.

a)



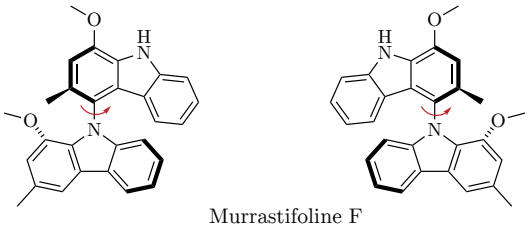
1,1′−Bi−2−naphthol

MolBar | 1.2.0 | ... | ... −28 −27 −25 −22 −22 12 27 42 45 67 79 83 84 107 111 138 169 201 216 249 282 295 345 473 507 723 1865 | −11 0 41

MolBar | 1.2.0 | ... | ... −28 −27 −25 −22 −22 12 27 42 45 67 79 83 84 107 111 138 169 201 216 249 282 295 345 473 507 723 1865 | −41 0 11

InChI=1S/C20H14O2/c21−17−11−9−13−5−1−3−7 15(13)19(17)20−16−8−4−2−6−14(16)10−12−18(20)22/h1−12, 21−22H

b)



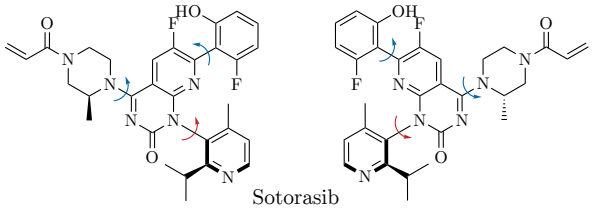
Murrastifoline F

MolBar | 1.2.0 | ... | ... 49 55 66 73 80 83 95 102 120 144 170 178 203 228 267 282 292 336 421 424 559 602 661 667 851 2244 | −20 0 0 0 0 0 50

MolBar | 1.2.0 | ... | ... 49 55 66 73 80 83 95 102 120 144 170 178 203 228 267 282 292 336 421 424 559 602 661 667 851 2244 | −50 0 0 0 0 0 20

InChI=1S/C28H24N2O2/c1−16−13−20−18−9−6−8−12−22(18) 30(28(20)24(14−16)32−4)27−17(2)15−23(31−3)26−25(27)19−10 −5−7−11−21(19)29−26/h5−15,29H,1−4H3

c)



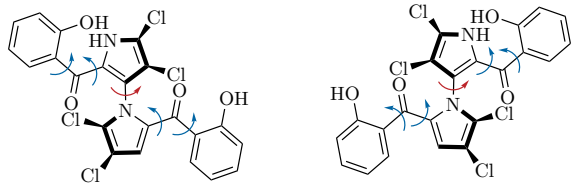
Sotorasib

MolBar | 1.2.0 | ... | ... 188 201 211 237 263 274 302 354 361 378 415 445 488 625 660 832 1063 1408 1818 | −150 −10 0 0 0 0 0 0 0 16 84

MolBar | 1.2.0 | ... | ... 188 201 211 237 263 274 302 354 361 378 415 445 488 625 660 832 1063 1408 1818 | −62 −11 0 0 0 0 0 0 0 16 177

InChI=1S/C30H30F2N6O3/c1−6−23(40)36−12−13−37(18(5)1 5−36)28−19−14−21(32)26(24−20(31)8−7−9−22(24)39)34−29(19)38(30(41)35−28)27−17(4)10−11−33−25(27)16(2)3/h6−11,14 ,16,18,39H,1,12−13,15H2,2−5H3/t18−/m0/s1

d)



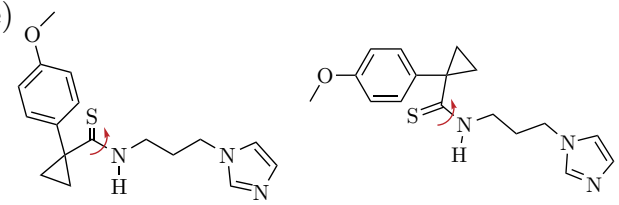
Marinopyrrole A

MolBar | 1.2.0 | ... | ... 82 83 85 99 118 137 157 158 187 219 227 246 270 282 317 344 344 416 418 619 1043 1066 1591 | −33 0 0 0 0 0 103

MolBar | 1.2.0 | ... | ... 82 83 85 99 118 137 157 158 187 219 227 246 270 282 317 344 344 416 418 619 1043 1066 1591 | −103 0 0 0 0 0 33

InChI=1S/C22H12Cl4N2O4/c23−12−9−13(19(31)10−5−1−3− 7−14(10)29)28(22(12)26)18−16(24)21(25)27−17(18)20(32)11−6 −2−4−8−15(11)30/h1−9,27,29−30H

e)



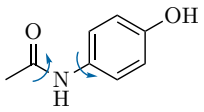
CHEMBL383477

MolBar | 1.2.0 | ... | ... −21 −3 7 9 9 9 9 9 12 21 38 55 63 66 84 99 105 143 175 191 213 234 276 302 534 569 780 912 1065 1232 | 0 0 0 0 0 0 0 0 0

MolBar | 1.2.0 | ... | ... −21 −4 7 9 9 9 9 9 15 28 38 54 60 65 83 104 105 138 175 186 213 234 276 310 535 569 776 912 1063 1234 | 0 0 0 0 0 0 0 0 0

InChI=1S/C17H21N3OS/c1−21−15−5−3−14(4−6−15)17(7−8 −17)16(22)19−9−2−11−20−12−10−18−13−20/h3−6,10,12−13 H,2,7−9,11H2,1H3,(H,19,22)

f)



Paracetamol

MolBar | 1.2.0 | ... | −141 −70 −60 −50 −32 −29 −26 −5 9 9 41 72 105 121 168 263 276 385 710 1096 | 0 0 0 0 0

InChI=1S/C8H9NO2/c1−6(10)9−7−2−4−8(11)5−3−7/h2−5,1 1H,1H3,(H,9,10)

Figure 6.6: Molecules with and without rotationally hindered single bonds and their MolBar representations. Atropisomeric axes with a predicted rotational barrier above 20 kcal/mol are indicated in red, while those below the threshold are shown in blue. The MolBar representations display deviating eigenvalues, color-coded as follows: red signifies a change in relative configuration (topography), and orange indicates a change in absolute configuration. The MolBar representation effectively captures the stereochemical information of the atropisomeric axes, allowing clear differentiation between stereoisomers directly from 3D coordinates, without the need for quantum chemical calculations or additional input. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

6.3 Permutation Invariance Test

Registry and storage systems require a unique and unambiguous identifier for each molecular structure. This uniqueness depends on establishing a one-to-one correspondence between the identifier and the molecule itself. For a molecule with n_{atoms}, there are $n_{\text{atoms}}!$ possible permutations of its atoms. SMILES and InChI resolve this by applying a canonicalization process to ensure a consistent atomic numbering scheme.[33] Prior studies have verified that InChI demonstrates permutation invariance.[260]

To evaluate MolBar's robustness against changes in atomic order, a permutation test was conducted. The test utilized 10,000 molecules randomly selected from the QMugs dataset.[56] The atomic numbering of these molecules was randomly permuted five times for each molecule. MolBar identifiers were generated for each permutation and compared to the original identifiers. The average atomic displacement was calculated for each permutation as a measure of the permutation's extent. This involved determining the distance between an atom in the permuted molecule and the atom with the same index in the original molecule, then averaging these distances across all atoms in the permutation. Figure 6.7 illustrates the average atomic displacement for each permuted molecule. The x-axis indicates the permutation index, while the y-axis displays the average atomic displacement in red and the number of MolBar changes in blue. No changes in MolBar identifiers were detected across all permutations, confirming that MolBar remains invariant to atomic order.

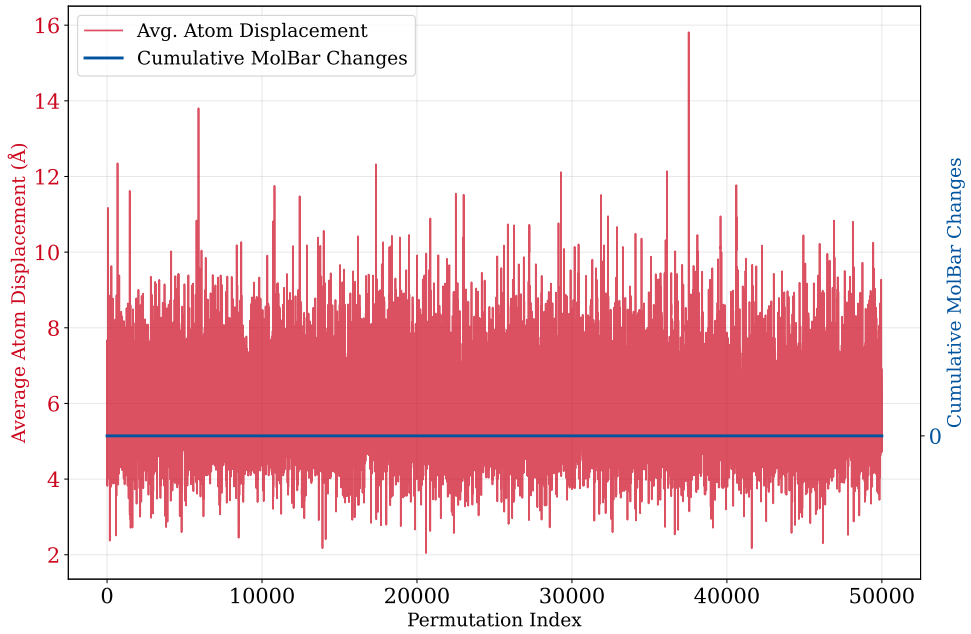


Figure 6.7: Permutation invariance test for MolBar. The average atomic displacement (red) is plotted against the permutation index, showing increasing displacement with each permutation. The number of MolBar changes (blue) remains zero across all permutations, indicating that MolBar is invariant to atomic order.

While the permutation invariance of eigenvalue spectra in MolBar may seem intrinsic, it is

essential that the entries in the various MolBar matrices also uphold this property. For example, maintaining an identical set of flexible, freely rotatable bonds is crucial. This ensures the molecule is consistently fragmented into matching sets of fragments, regardless of atomic order. Likewise, force field constraints must be defined in a permutation-invariant manner to ensure a unified representation.

6.4 Assessment of Computational Efficiency of MolBar

An evaluation of the computational efficiency associated with MolBar was conducted by measuring the time required to generate its identifier. This assessment utilized a randomly selected subset of 1000 molecules from the QMugs dataset.[56]

The measurements were performed using the `molbar` Python package, specifically version 1.2.0, employing the function `get_molbar_from_coordinates`. To ensure clarity in understanding how computational cost scales with molecular size, the timing results were averaged for molecules containing an identical number of atoms. For comparative purposes, the time required to generate InChI strings was also recorded. This was achieved using the `rdkit.Chem.MolToInchi` function from the `rdkit` package, version 2025.3.5.[86] It is important to note that InChI strings were generated from 3D coordinates. This process involved the use of the `rdkit.Chem.rdDetermineBonds.DetermineBonds` function, which, while not part of the core InChI implementation, is essential for converting 3D structures into molecular graphs with defined bond orders. This step was included to ensure a fair and accurate comparison. Given that MolBar is primarily intended for applications in chemical space exploration or PES analysis, the timings for single-point energy calculations were also considered. These included calculations using GFN2-xTB via the `tblite` package, version 0.5.0,[46,261] and UMA-s-1 MLIP via the `fairchem-core` package, version 2.11.0.[65] All timings were conducted on Apple Silicon M1 Pro hardware.

Figure 6.8 illustrates the computational cost of MolBar in comparison to InChI generation and single-point energy calculations using GFN2-xTB and UMA-s-1. The cost is plotted as a function of the number of atoms, including hydrogens, within the molecule. The computational cost of MolBar exhibits a near-linear relationship with the number of atoms in the molecule, characterized by a $\mathcal{O}(n_{\text{atoms}}^{0.93})$ scaling. In contrast, InChI generation demonstrates a slightly sub-linear scaling, with a $\mathcal{O}(n_{\text{atoms}}^{0.61})$ relationship. Despite this, InChI generation remains approximately two orders of magnitude faster than MolBar across the examined molecular sizes. However, a few outliers were observed in the InChI generation process. Further investigation revealed that these outliers were primarily due to the bond order assignment step, which is not inherently part of the InChI generation but is necessary for converting 3D coordinates into a molecular graph with defined bond orders. This step is likely to be the rate-determining factor for InChI generation from 3D structures, and alternative implementations for this conversion could yield different computational costs. In contrast, MolBar did not exhibit significant outliers within this dataset. Although MolBar is more computationally demanding than InChI, it remains sufficiently efficient for its intended applications in chemical space exploration. For molecules containing up to approximately 80 atoms, the computational cost of MolBar is comparable to that of a single-point energy calculation using GFN2-xTB. Given the higher scaling exponent of GFN2-xTB ($\mathcal{O}(n_{\text{atoms}}^{1.81})$), MolBar becomes relatively more cost-effective for larger molecules. When compared to UMA-s-1, MolBar consistently demonstrates

lower computational costs across the examined size range. While the scaling exponent of UMA-s-1 ($\mathcal{O}(n_{\text{atoms}}^{1.24})$) is lower than that of GFN2-xTB, it remains more expensive than MolBar ($\mathcal{O}(n_{\text{atoms}}^{0.93})$) for the molecular sizes considered. In practical applications, MolBar is generally computed following a geometry optimization, conformer sampling, or MD simulation, each of which involves multiple energy and gradient evaluations to yield relevant 3D structures. As a result, the computational cost of MolBar becomes negligible when set against the total cost of these workflows.

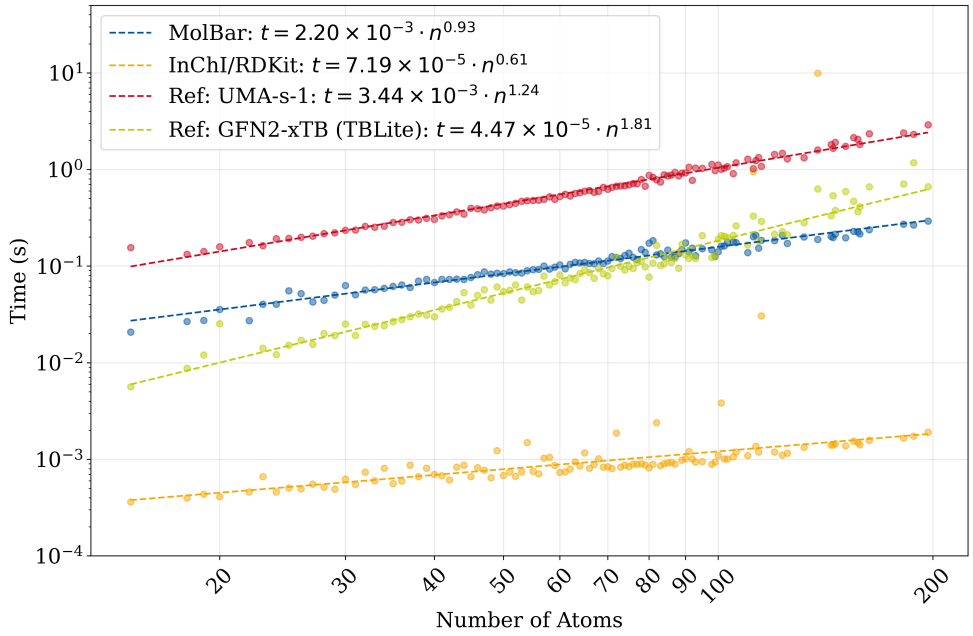
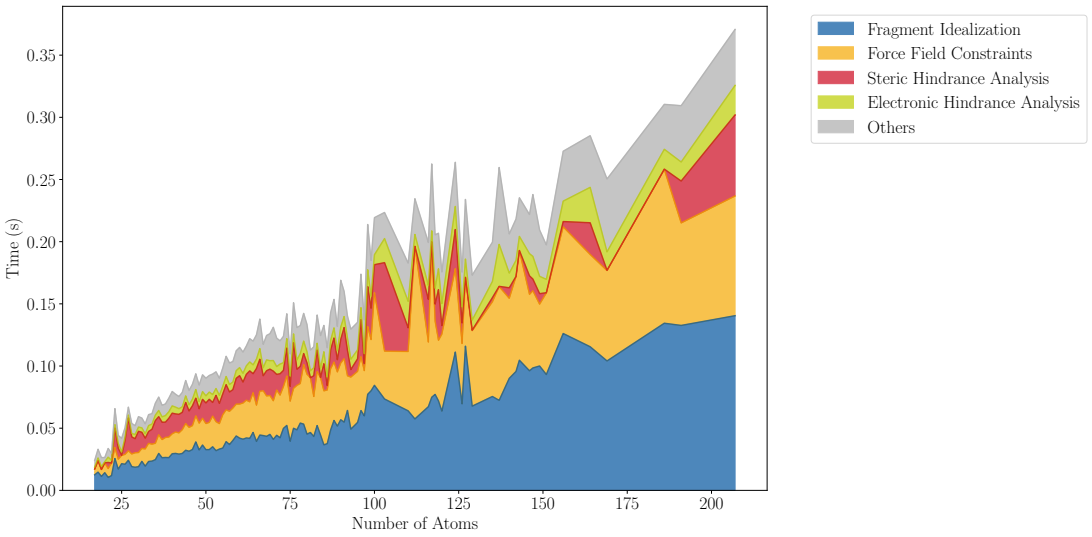


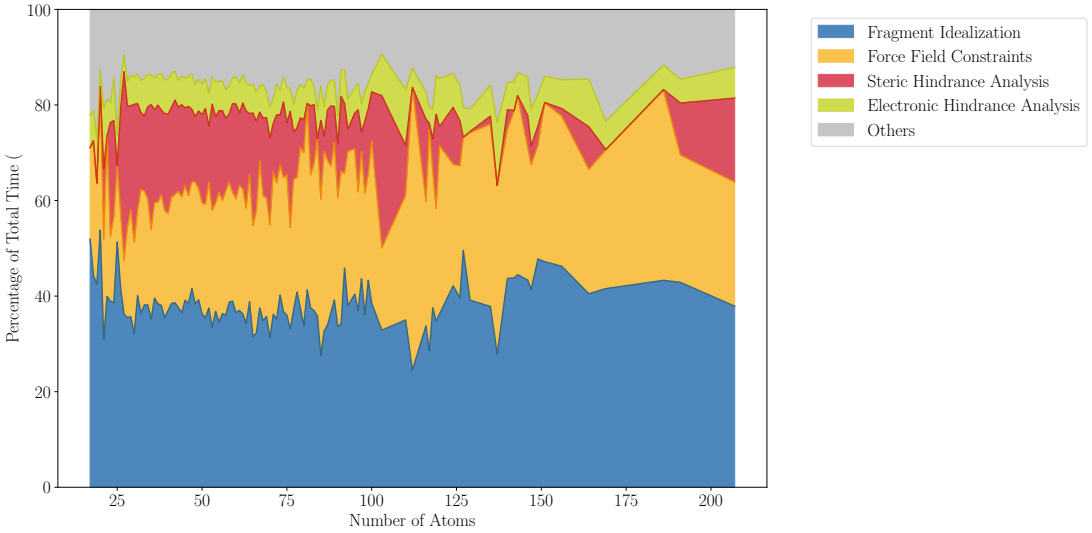
Figure 6.8: Computational cost comparison of MolBar, InChI, GFN2-xTB, and UMA-s-1 for the QMugs dataset. The time required to calculate the MolBar identifier is shown for molecules grouped by atom count.

It should be emphasized that the time needed to calculate the MolBar identifier can differ substantially. The number of rings, molecular rigidity, and the presence of freely rotatable bonds can all contribute to variations in calculation times, even among molecules with identical atom counts. Therefore, the reported timings are purely indicative and should be approached with care, given their strong dependence on the specific molecular structure. For instance, the photoactive yellow protein, which contains 1931 atoms, required an average of approximately 69 seconds to compute the MolBar identifier on the same hardware across five runs. This duration is significantly longer than the value predicted by the fitted trend in Figure 6.8 (predicted: 1.98 seconds), illustrating how molecular complexity influences computational cost.

To further elucidate the distribution of computational costs within MolBar, a detailed timing breakdown of its primary components was conducted. The five most computationally intensive components were identified as: fragment idealization, geometry constraints, steric hindrance, topology determination (including connectivity and electronic hindrance), and chirality determination. Figure 6.9 presents the absolute and relative timing breakdown of these components as a function of molecular size.



(a) Absolute timing breakdown



(b) Relative timing breakdown

Figure 6.9: Breakdown of the computational cost of MolBar into its main components: fragment idealization, geometry constraints, steric hindrance, topology determination (connectivity and electronic hindrance), and chirality determination. The absolute (a) and relative (b) timings of each component are shown as a function of the number of atoms in the molecule. Fragment idealization emerges as the most computationally demanding step, accounting for approximately 45% of the total cost across the examined size range.

Fragment idealization represents the most computationally intensive step in the MolBar process, accounting for roughly 45% of the total computational cost across the examined molecular sizes. This step involves the application of force field constraints and the execution of a geometry optimization for each fragment to obtain its idealized geometry (see Section 5.5). While the MolBarOpt proves to be more robust and computationally efficient than other analyzed optimizers, such as SciPy's Newton-CG[135] and CREST's ANCOpt Python clone,[234] it still constitutes a significant portion of the overall computational cost (see Section 5.5). This

underscores the importance of developing efficient geometry optimization algorithms within the context of MolBar calculations.

The computational demand associated with geometry constraints may initially seem surprising. This is primarily due to the Kabsch-based alignment procedure,[240] which is applied to each atom's neighborhood and all its possible permutations. This step requires the calculation of the optimal rotation matrix to align the actual and idealized local geometries, enabling the assignment of the correct set of geometry constraints. Given the involvement of multiple matrix operations, this process can be computationally intensive, particularly for larger molecules with a high number of atoms.

6.5 Application of MolBar as a Molecular Representation in Databases and Database Deduplication

Maintaining a single, authoritative record per molecule is essential for database reliability and long-term functionality. Deduplication ensures data quality by eliminating redundancies, optimizing storage efficiency, and preventing the propagation of outdated or incorrect information. A key use case for molecular identifiers is enabling a unique representation of molecules in databases. In 2D cheminformatics databases based on Ctab files, molecules can be represented in different ways, such as various resonance structures. In 3D quantum chemistry databases, Cartesian coordinates for the same molecule can vary due to translation, rotation, conformation, or theory level. A molecular identifier allows for a one-to-one mapping between molecular identity and the Ctab or XYZ file. Thus, molecular identifiers are used for deduplication tasks.

The performance of MolBar, InChI, and SMILES is evaluated for deduplicating quantum chemistry datasets containing 3D structures. Two common datasets of quantum chemistry are evaluated in the following: QMugs[56] dataset containing 665,911 organic drug-like molecules and the tmQM[57] dataset with 107,998 transition metal complexes, covering all 30 metals in the 3d, 4d, and 5d series, and over 30,000 unique ligands. For QMugs, MolBar is compared with FixedH-InChI generated from the SDfile and `RDKit` version 2025.9.4.[86] Fixed-H InChI is used here as it does not normalize different tautomer representations, and distinguishes between them as MolBar does. The tmQM dataset comes with machine-readable SMILES representations, which are used to benchmark MolBar.[122]

Table 6.1 provides an overview of the duplicate detection results for MolBar and FixedH-InChI on the QMugs dataset. Out of 665,911 molecules, 657,121 (98.68%) are unique and appear only once. Approximately 1.26% of the molecules, totaling 8,403, were flagged as duplicates by both MolBar and InChI. A duplicate is defined as an identifier that has appeared at least once before in the dataset. The first occurrence of an identifier is not considered a duplicate, while subsequent occurrences are classified as such.

Table 6.1: Overview of duplicate detection in the QMugs dataset using MolBar and FixedH-InChI identifiers. Duplicates are defined as identifiers that have appeared at least once before in the dataset, with the first occurrence not considered a duplicate. When both methods identify the same molecule as a duplicate, it is counted as a duplicate for both methods. When only one method identifies a molecule as a duplicate, it is counted as a duplicate for that method only.

Metric	Count	Percentage
Duplicates (both methods)	8,403	1.26%
Duplicates (MolBar only)	4	<0.01%
Duplicates (InChI only)	383	0.06%
Unique molecules (both methods)	657,121	98.68%
Total	665,911	100.0%

Four molecules are identified as MolBar-only duplicates, and these cases do not represent errors by MolBar. They originate from the parser converting 3D coordinates to InChI, rather than being an issue with InChI itself. As such, these examples are not addressed further here.

Approximately 383 additional molecules are classified as duplicates exclusively by InChI, whereas MolBar identifies distinctions. These distinctions are organized into categories in Table 6.2. Examples for each category are illustrated in Figure 6.10. The duplicate FixedH-InChI is displayed for both structures. The differing MolBar identifiers are also shown, with the eigenvalues where MolBar detects changes highlighted. Changes are color-coded by the affected component: blue indicates topology, red indicates topography, and orange indicates chirality. For clarity, only relevant parts of the MolBar identifiers are included.

Table 6.2: Categorization of molecular differences in the 356 duplicate groups containing 383 molecules identified by InChI but not by MolBar. These cases represent distinct molecular structures that share the same InChI identifier but have different MolBar representations.

Category	Groups	Percentage
Atropisomer	58	16.3%
Amide atropisomers	227	63.8%
Non-central chirality	17	4.8%
Rigidity analysis	4	1.1%
Different connectivity	10	2.8%
Different topography	40	11.2%
Total	356	100.0%

Six distinct categories were identified, with the first three representing cases that MolBar correctly identifies as distinct molecules, accounting for 84.8% of the instances. The remaining three categories correspond to cases where MolBar makes errors (15.2%).

The first example involves atropisomers (58 groups, 16.3%), such as CHEMBL3672366 and CHEMBL3672340, which MolBar identifies as atropisomers. Both molecules are quinolinone derivatives featuring a meta-substituted fluoraryl substituent, whose rotation relative to the quinolinone backbone is restricted. The computational barrier discussed in Section 5.4 exceeds 20 kcal/mol, supporting MolBar's prediction. MolBar classifies both structures as diastereomers due to the presence of another stereocenter. InChI, in its current version, cannot represent axial chirality or atropisomerism, as previously noted in Section 5.4.

Figure 6.10b illustrates an example of amide isomerization involving a thioamide (227 groups, 63.8%) in CHEMBL39999 and CHEMBL543611. MolBar predicts an amide bond with a bond order above 1.25, resulting in increased rotational barriers and distinct *cis*/*trans* isomers. InChI interprets this bond as a single bond without associated stereoinformation.

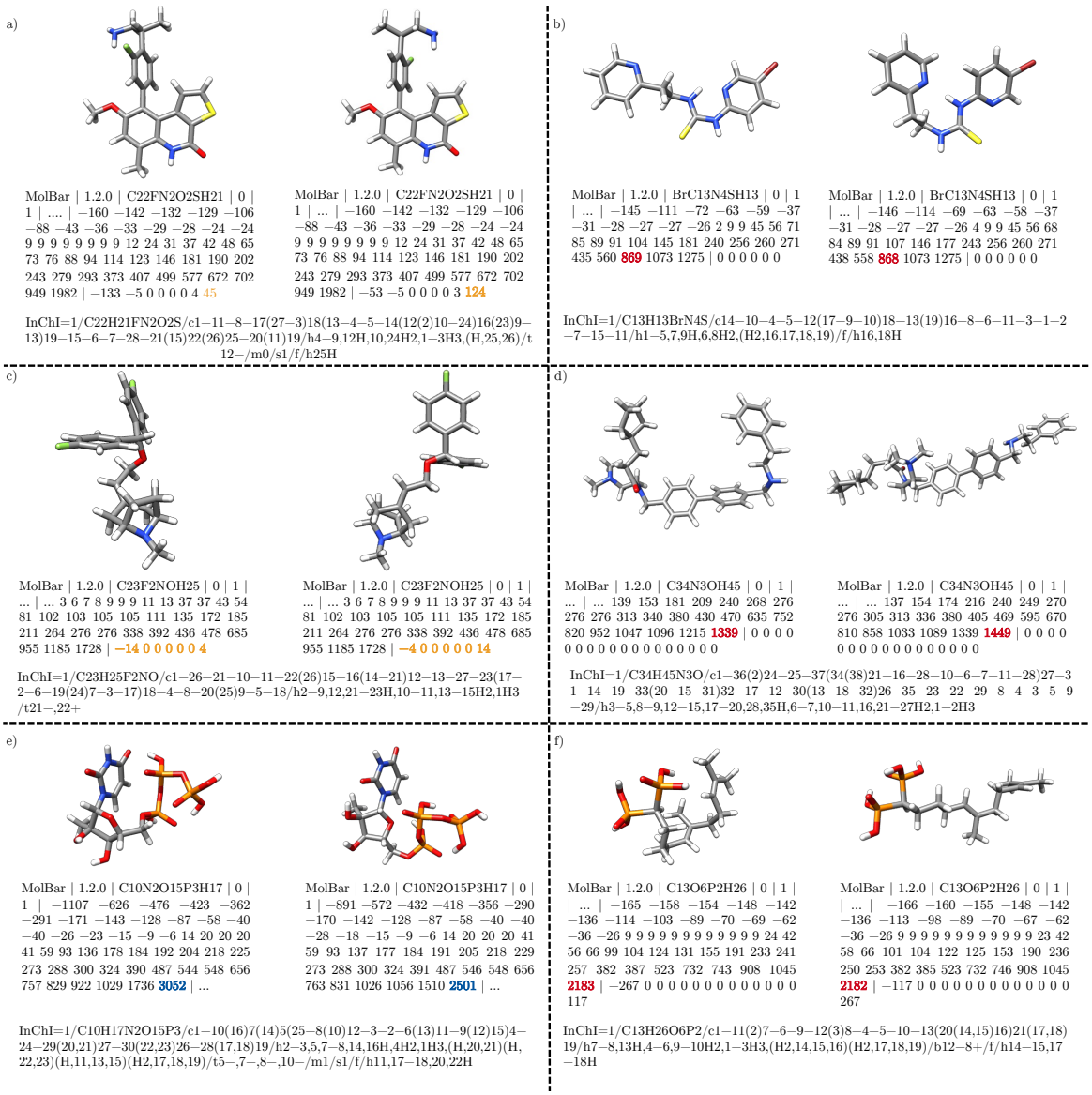


Figure 6.10: Representative molecular pairs from the QMugs dataset that share identical InChI identifiers but exhibit distinct MolBar representations. The duplicate FixedH-InChI is shown for both structures. The differing MolBar identifiers are displayed, with eigenvalues highlighting where MolBar detects changes. Color-coding indicates the affected component: blue for topology, red for topography, and orange for chirality. Only the relevant parts of the MolBar identifiers are presented for clarity. InChI classifies these structures as duplicates, while MolBar accurately differentiates them based on: (a) biaryl atropisomerism, (b) amide atropisomerism, and (c) axial chirality. The cases in (d), (e), and (f) illustrate instances where MolBar makes errors: (d) different molecular fragmentation for the same molecule, (e) connectivity differences, and (f) topographical variations due to incorrect connectivity assignment.

Figure 6.10c demonstrates axial chirality for CHEMBL129193 and CHEMBL1927861 (17 groups, 4.8%), where the carbon cage with an attached double bond creates a chiral object. The double bond itself does not exhibit *cis*/*trans* isomerism, as both double bond substituents from the cage are equivalent. Chirality arises from the interaction between the cage's curvature and the double bond. MolBar treats the cage and double bond as a single fragment, classifying both

structures as enantiomers. InChI cannot describe axial chirality, rendering both identifiers equivalent.

Figure 6.10d presents cases where InChI is correct. Geometric filtering of bonds with high WBO or sterically hindered bonds (where dihedrals are compared to thresholds, as non-planar bonds lack π-overlap and sterically hindered bonds cannot be flat) leads to different fragmentation of conformers of the same molecule. These cases are rare (4 groups, 1.1%) and are not discussed further.

The final two examples highlight instances where MolBar makes mistakes. These include cases where different conformers are evaluated as having distinct connectivity (10 groups, 2.8%) or different topography (40 groups, 11.2%). Both issues stem from the same phenomenon: when two phosphane groups are near each other, MolBar's connectivity analysis often detects a bond between the phosphorus atoms. In the first scenario, one conformer lacks this bond, while the other includes it, resulting in different MolBar identifiers due to connectivity. In the second scenario, both conformers exhibit the same unreasonable P–P bond. Due to insufficient constraints in fragment idealization for such geometries, both structures are idealized into slightly different geometries. This suggests that MolBar's radii for phosphorus require adjustment. Using valences to limit the maximum number of neighbors for phosphorus is impractical, as this would restrict MolBar to the chemistry of the main groups, where valences only apply. Further, it would be unsuitable for transition states when exploring chemical space, as valences do not apply here either.

The results demonstrate that MolBar provides superior discrimination for stereochemical and conformational differences, particularly for atropisomers and topographical variations. The majority of InChI-only duplicates (63.8%) were amide atropisomers, which InChI does not distinguish but MolBar correctly identifies as distinct structures.

To go from organic to inorganic chemistry, Table 6.3 summarizes duplicate detection results for the tmQM dataset using MolBar and SMILES identifiers. Among 107,998 transition metal complexes, 94,098 (87.13%) are classified as unique by both approaches. A combined total of 2,624 molecules (2.43%) are marked as duplicates through MolBar and SMILES.

Table 6.3: Overview of duplicate detection in the tmQM dataset using MolBar and SMILES identifiers.

Metric	Count	Percentage
Duplicates (both methods)	2,624	2.43%
Duplicates (MolBar only)	229	0.21%
Duplicates (SMILES only)	11,047	10.23%
Unique molecules (both methods)	94,098	87.13%
Total	107,998	100.0%

MolBar exclusively identifies 229 molecules (0.21%) as duplicates, primarily due to its recog-

nition of identical connectivity where SMILES interprets a ligand as detached. Such discrepancies stem from differing thresholds between the two methods. A further 11,047 molecules (10.23%) are flagged as duplicates solely by SMILES, while MolBar treats them as distinct entities. Although some SMILES extensions support basic metal complex stereochemistry,[208] the dataset's SMILES encoding does not include stereochemical details. Given MolBar's full stereochemical representation, it demonstrates greater sensitivity and identifies fewer duplicates. Figure 6.11 illustrates example molecular pairs from the tmQM dataset, each sharing identical SMILES identifiers but differing in their MolBar representations. Specifically, MolBar effectively differentiates (a) the axial chirality of a rhenium complex and (b) the *cis-trans* isomerism of a platinum complex. The duplicate SMILES is displayed for both structures. The differing MolBar identifiers are also shown, with the eigenvalues where MolBar detects changes highlighted. Changes are color-coded by the affected component: blue indicates topology, red indicates topography, and orange indicates chirality. For clarity, only relevant parts of the MolBar identifiers are included.

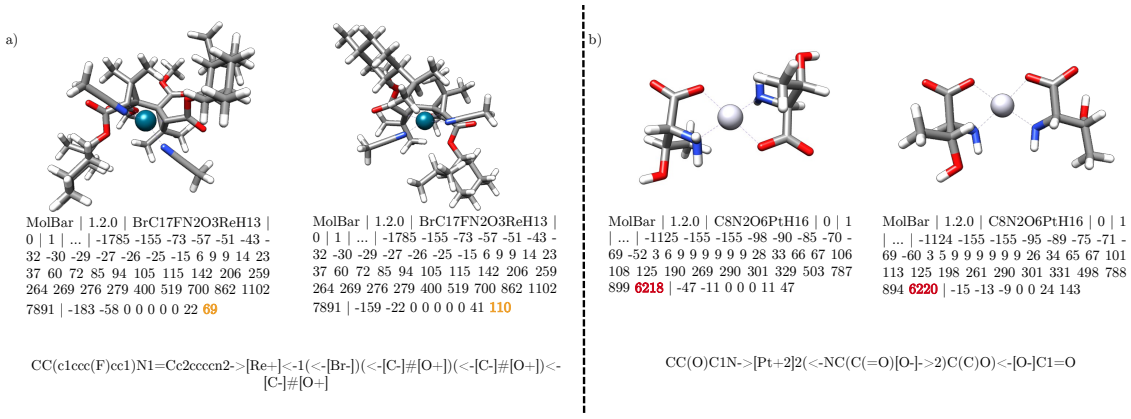


Figure 6.11: Representative molecular pairs from the tmQM dataset that share identical SMILES identifiers but possess distinct MolBar representations. The duplicate SMILES is displayed for both structures. The differing MolBar identifiers are also shown, with the eigenvalues where MolBar detects changes highlighted. Changes are color-coded by the affected component: blue indicates topology, red indicates topography, and orange indicates chirality. For clarity, only relevant parts of the MolBar identifiers are included. While SMILES treats these structures as duplicates, MolBar correctly distinguishes (a) axial chirality and (b) *cis-trans* isomerism.

Across both QMugs and tmQM, MolBar yields fewer false duplicates than identifier-based baselines because it applies to the full periodic table, covering both organic and transition-metal chemistry. It captures key stereochemical effects, including axial chirality, *cis-trans* isomerism in metal complexes, atropisomerism, and configurationally stable amides. Most entries flagged as duplicates by InChI or SMILES correspond to genuinely distinct structures that MolBar separates correctly, and the remaining mismatches can be traced to specific connectivity and radii limitations. Overall, these results support MolBar as a reliable deduplication identifier for 3D structure databases. Compared to existing identifiers, MolBar removes duplicates with the highest precision on common quantum-chemical datasets such as QMugs and tmQM. It is the only identifier that can automatically encode these stereochemical cases directly from

3D coordinates without manual input. MolBar thus stands out as a highly effective tool for deduplication in cheminformatics databases, especially those containing complex stereochemical information.

6.6 Application of MolBar in Chemical Space Exploration

Beyond its application in molecular databases, MolBar facilitates chemical space exploration by linking 3D coordinates to molecular identity.

This capability proves especially valuable for tools like GOAT,[47] CREST,[48] and TTConf,[49] which generate new 3D structures through black box methods. MolBar allows for the identification of conformers that share the same molecular identity as the input by assigning them identical MolBar values. For reaction exploration and the generation of new molecules, distinct MolBar values indicate different molecular identities. This applies to approaches such as the nanoreactor method by Martinez and coworkers,[50] Chemoton by M. Reiher and coworkers,[51] the QCxMS2 method by J. Gorges and S. Grimme,[52] and aRST by Y. Chen and C. Bannwarth.[53] Since MolBar largely disregards chemical rules and valences, it can be applied to completely unknown and arbitrary structures.

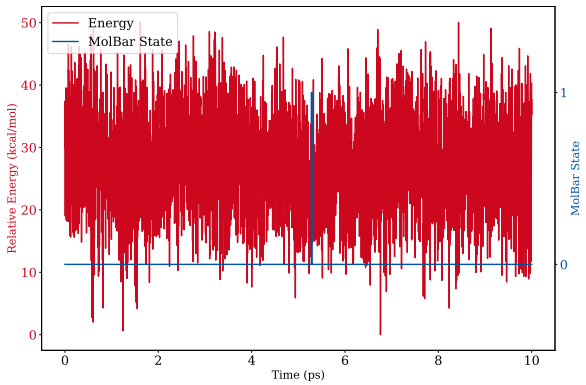
A common requirement across these methods is the need to compare generated structures with the input to determine whether the molecular identity remains unchanged or has been altered. Given the volume of output, manual verification of thousands of 3D structures is impractical. MolBar automates this process, ensuring results align with the intended objectives of each tool.

The following sections explore the application of MolBar in these contexts. First, the structures from an MD simulation are analyzed to confirm the molecule's integrity throughout the simulation. Next, CREST conformer ensembles are examined to filter out only true conformers. Finally, a simplified implementation of the *ab initio* nanoreactor, along with an educational example, is analyzed. Here, MolBar is used for molecular bookkeeping and the identification of molecular reactions.

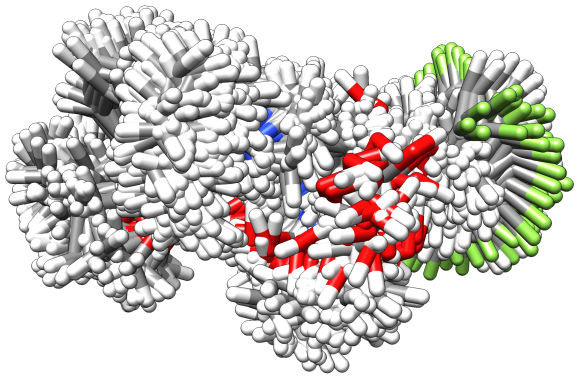
6.6.1 Molecular Dynamics Simulation

First, CHEMBL1209357[262] is analyzed for structural changes during a high-temperature MD simulation. The simulation used the `ase`[263] package in Python, with electronic structure calculations conducted using the GFN2-xTB semi-empirical tight-binding method via the TBLite library.[46,261] Before initiating MD simulations, the molecule's structure underwent geometric optimization using the BFGS[264] algorithm, applying a convergence threshold of 0.01 eV/Å for the maximum force. Systems were equilibrated at 1000 K for 1 ps (1000 steps) with the velocity Verlet integrator and a timestep of 1.0 fs. Initial velocities were drawn from a Maxwell-Boltzmann distribution corresponding to the target temperature. The production MD runs lasted 10 ps (10,000 steps) at 1000 K, maintaining the same integrator and timestep settings. MolBar values were calculated for all 10,000 structures generated during the trajectory using MolBar version 1.2.0.[189] Conformational states were categorized by assigning unique integer identifiers to each distinct MolBar encountered.

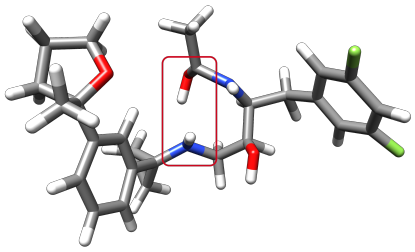
Figure 6.12a illustrates the MD simulation trajectory for CHEMBL1209357. The sampled structures during the simulation exhibited an average electronic energy approximately 30 kcal/mol above the optimized minimum. MolBar analysis revealed that only 0.2% (16 out of 10,000) of the structures produced distinct MolBar identifiers. This result highlights the conformational invariance of MolBar under the tested conditions. Despite the severe structural changes seen among 100 sampled structures in Figure 6.12b, MolBar's rigidity analysis and structure idealization allow for the description of the 3D structure of the molecule in the topography matrix without requiring conformational information. This also confirms that the derived force constraints in the geometry analysis are robust to conformational changes. All in all, almost all structures are described by the same MolBar identifier, confirming that the molecule's identity remains mostly unchanged throughout the simulation. Figures 6.12c and 6.12d show the rare cases where MolBar identifiers diverged. The former presents the initial structure, while the latter depicts a scenario where hydrogen bonding between the amine group and an sp^2-hybridized carbon atom caused a change in the MolBar identifier.



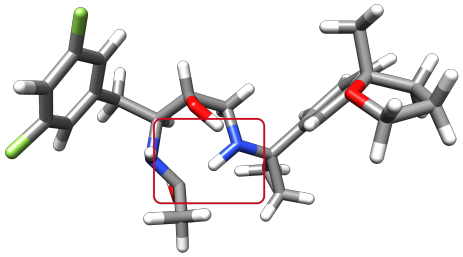
(a) MD trajectory of CHEMBL1209357.



(b) 100 sampled structures of MD simulation



(c) Initial molecular structure.



(d) Structure with a distinct MolBar identifier due to hydrogen bonding.

Figure 6.12: MD simulation of CHEMBL1209357, demonstrating MolBar's conformational invariance. Only 0.2% of structures showed altered identifiers, attributable to specific hydrogen bonding interactions.

6.6.2 Conformer Sampling

A primary application of MolBar is in the analysis of conformer ensembles generated by tools like CREST.[234] MolBar was used to filter CREST[234] ensembles to only include true conformers

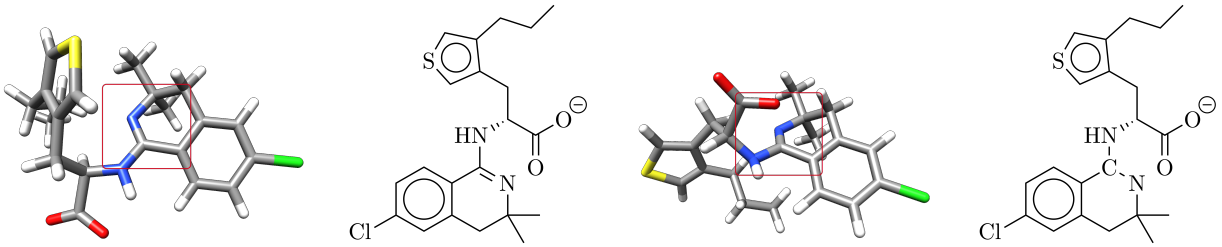
and remove structures where structural changes occurred. The BACE[225] dataset, which includes 1511 molecules and their CREST[234] conformer ensembles, provided the basis for this investigation. A random subset of 100 ensembles, totaling 18,678 conformers, was evaluated using MolBar and compared with canonical SMILES representations generated via `OpenBabel`[85] (version 3.1.0).

Table 6.4: Analysis of conformer ensembles from the BACE dataset comparing MolBar and canonical SMILES representations across 100 randomly selected ensembles.

Category	Count	Percentage
Conformational variation only	18516	99.1%
SMILES artifact (MolBar unchanged)	104	0.6%
Amide isomerization	9	<0.1%
Molecular transformation detected	49	0.3%

Four categories emerged from comparing MolBar identifiers and canonical SMILES representations: conformational variation only (identical MolBar and SMILES), SMILES artifacts (identical MolBar but different SMILES), amide isomerization (a specific conformational change identified by MolBar), and molecular transformations (different MolBar identifiers). The majority of conformers (99.1%) displayed only conformational variation, confirmed by identical MolBar identifiers and SMILES representations. A small fraction (0.6%) showed discrepancies in SMILES representations while retaining consistent MolBar identifiers. This discrepancy arises from bond order assignment based on 3D coordinates, where one structure may feature a double bond unrecognized in another conformer due to dihedral angle-based filtering of implausible double bonds. Figure 6.13 provides an example of this phenomenon, where the initial structure contains a double bond absent in the conformer. In such cases, the SMILES representation of the conformer fails to satisfy valences, as the missing double bond is not compensated by additional hydrogens. This issue depends on the bond order assignment implementation and does not reflect a fundamental limitation of the MolBar identifier. While alternative tools might yield different results, `OpenBabel` was selected for this application due to its handling of unsatisfied valences, which is common in conformer ensembles with fluctuating bond orders.

In 9 cases (<0.1%), amide isomerization was detected, a specific conformational change identifiable by MolBar due to alterations in the local bonding environment. MolBar classifies a bond as rotationally hindered if the WBO exceeds 1.25, roughly corresponding to a barrier of 20 kcal/mol, which is typical for amide bonds. SMILES representations, relying on integer bond orders, treat this bond as a true single bond without stereoinformation. Figure 6.14 illustrates an example of amide isomerization, where the initial structure contains a trans amide bond and the conformer exhibits a cis amide bond. The computed rotational barrier at UMA[65]//GFN2-xTB[46] level of theory with the protocol from Section 5.4 equals 20.0 kcal/mol,



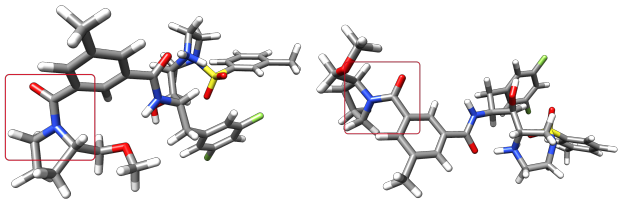
(a) Initial structure.

(b) SMILES representation of the initial structure.

(c) Conformer structure.

(d) SMILES representation of the conformer structure.

Figure 6.13: Example of a conformer where the MolBar identifier remains unchanged, indicating conformational variation only, while the canonical SMILES representation differs due to a double bond recognized only in the initial structure. This results from bond order depiction using dihedral angles to filter out implausible double bonds.



(a) Initial structure.

(b) Conformer.

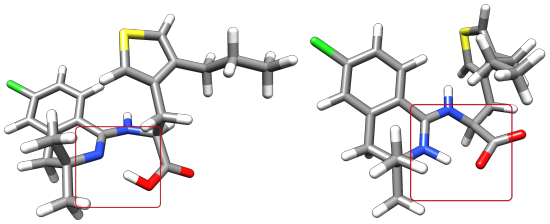
Figure 6.14: Example of amide isomerization detected by MolBar, where the initial structure contains a trans amide bond and the conformer exhibits a cis amide bond. The computed rotational barrier for this transformation is 20.0 kcal/mol, surpassing the threshold for rotationally hindered bonds in MolBar.

matching MolBar's threshold for rotationally hindered bonds, thus leading to a change in the MolBar identifier.

A negligible number of cases (0.3%) involved molecular transformations, where both the MolBar identifier and SMILES representation changed. This indicates that the conformer represents a different molecule rather than a distinct conformation of the same molecule. A common example is tautomerization, where a proton shift alters the bonding pattern and results in a different MolBar identifier. Figure 6.15 shows an example of a molecular transformation, where the initial structure contains a hydroxyl group and the conformer features a carbonyl group due to tautomerization.

6.6.3 Reaction Discovery

MolBar also finds application in reaction discovery, where it functions as a tool for molecular bookkeeping and reaction identification. Numerous tools exist for reaction discovery, with the *ab initio* nanoreactor developed by Martinez and coworkers[50] serving as a notable example. This nanoreactor performs *ab initio* MD simulations of freely reacting molecules, generating new molecules and uncovering novel reaction mechanisms without relying on predefined heuristics or reaction coordinates, but instead using quantum mechanical and classical



(a) Initial structure. **(b)** Conformer structure.

Figure 6.15: Example of a molecular transformation detected by MolBar, where the initial structure contains a hydroxyl group and the conformer shows a hydrogen shift resulting in an amine group. This transformation leads to a different bonding pattern and thus a different MolBar identifier, indicating a distinct molecule rather than a conformational variant.

mechanical equations to describe molecular behavior. In this approach, a set of molecules is confined within a virtual sphere, and virtual pistons periodically push the molecules toward the sphere's center to accelerate reaction events, which are otherwise rare under normal conditions. The original implementation employed an approximate Hartree–Fock method for the MD simulations. The Urey–Miller experiment, first conducted in the 1950s to replicate early Earth conditions and explore the chemical origins of life, was simulated using this nanoreactor.

In the original experiment, a gas mixture of hydrogen, ammonia, methane, carbon monoxide, and water was exposed to electrical sparks, mimicking lightning, which resulted in the formation of organic compounds, including amino acids. The nanoreactor simulation successfully replicated the formation of amino acids and other organic molecules, demonstrating its ability to investigate complex chemical reactions without predefined pathways.[50] For educational purposes, a simplified version of the *ab initio* nanoreactor is presented here, with MolBar directly integrated for molecular bookkeeping and reaction identification, as outlined in Algorithm 4. The MD simulations are conducted using `ase`,[263] employing Langevin dynamics with the velocity Verlet integrator, a timestep of 0.5 fs, and a friction coefficient of 0.2 fs^{-1} at 300 K. The simulations utilize the GFN2-xTB[46] method through the TBLite[261] library, setting the electronic temperature to 1000 K and assuming a multiplicity of 1, while incorporating the ALPB[222] solvent model for water to account for solvation effects.

The simulation begins with the random placement of O_2, H_2, H_2O, 2 CO, and 2 NH_3 within a sphere of 6.0 Å radius. To expedite reaction events for this educational demonstration, molecules are permitted to overlap when randomly positioned inside the sphere. The first 10 steps of the simulation are excluded to allow the system to equilibrate. Following each MD step, a boundary handling procedure ensures that atoms remaining outside the sphere are returned to its interior. Elastic collisions are assumed at the boundary, where the radial momentum component is inverted to simulate a rebound into the sphere, while the tangential component remains unaltered. Additionally, a piston force is applied periodically for 50 steps every 200 steps, using a harmonic force to push atoms toward the sphere's center and fur-

ther accelerate reaction events. After each MD step, MolBar's connectivity analysis identifies separate molecules (see Section 5.2). A MolBar identifier is then calculated and recorded for each molecule in a history log, enabling the tracking of molecular formation and transformation throughout the simulation. This approach facilitates the detection of reaction events by observing changes in MolBar identifiers over time, allowing the construction of a reaction network based on the observed molecular transformations. Each structure, along with its MolBar identifier, is optimized at the GFN2-xTB[46] level of theory to assess its stability. If the MolBar identifier changes after optimization, the structure is marked as an unstable transition state.

Algorithm 4 Simplified *ab initio* Nanoreactor with MolBar bookkeeping

Parameters:
$steps \leftarrow 100000$
$period \leftarrow 200$
$radius \leftarrow 6.0$
$k \leftarrow 1000.0$

Initialize atomic positions $\mathbf{r}_i$, velocities $\mathbf{v}_i$

for $step \leftarrow 1$ **to** $steps$ **do**
    MD Step:
        $\mathbf{F} \leftarrow \textsc{ComputeForces}(\mathbf{r})$
        $(\mathbf{r}, \mathbf{v}) \leftarrow \textsc{LangevinStep}(\mathbf{r}, \mathbf{v}, \mathbf{F})$

    Boundary Handling:
    for each atom i **do**
        $d \leftarrow \|\mathbf{r}_i\|$
        if $d > radius$ **then**
            $\hat{r}_i \leftarrow \mathbf{r}_i / d$
            $\mathbf{r}_i \leftarrow radius \cdot \hat{r}_i$
            $\mathbf{v}_i \leftarrow \mathbf{v}_i - 2(\mathbf{v}_i \cdot \hat{r}_i)\hat{r}_i$
        end if
    end for

    Piston Force Application:
    if $step \bmod period < 50$ **then** ▷ Compression phase
        for each atom i **do**
            $\hat{r}_i \leftarrow \mathbf{r}_i / \|\mathbf{r}_i\|$
            $\mathbf{F}_i \leftarrow \mathbf{F}_i - k \cdot \hat{r}_i$
        end for
    end if

    Reaction Tracking:
    $\mathcal{C} \leftarrow \textsc{get_connected_components}(\mathbf{r})$
    $\mathcal{M} \leftarrow \textsc{get_molbar_from_coordinates}(\mathcal{C})$
    $\textsc{UpdateHistory}(\mathcal{M})$
end for

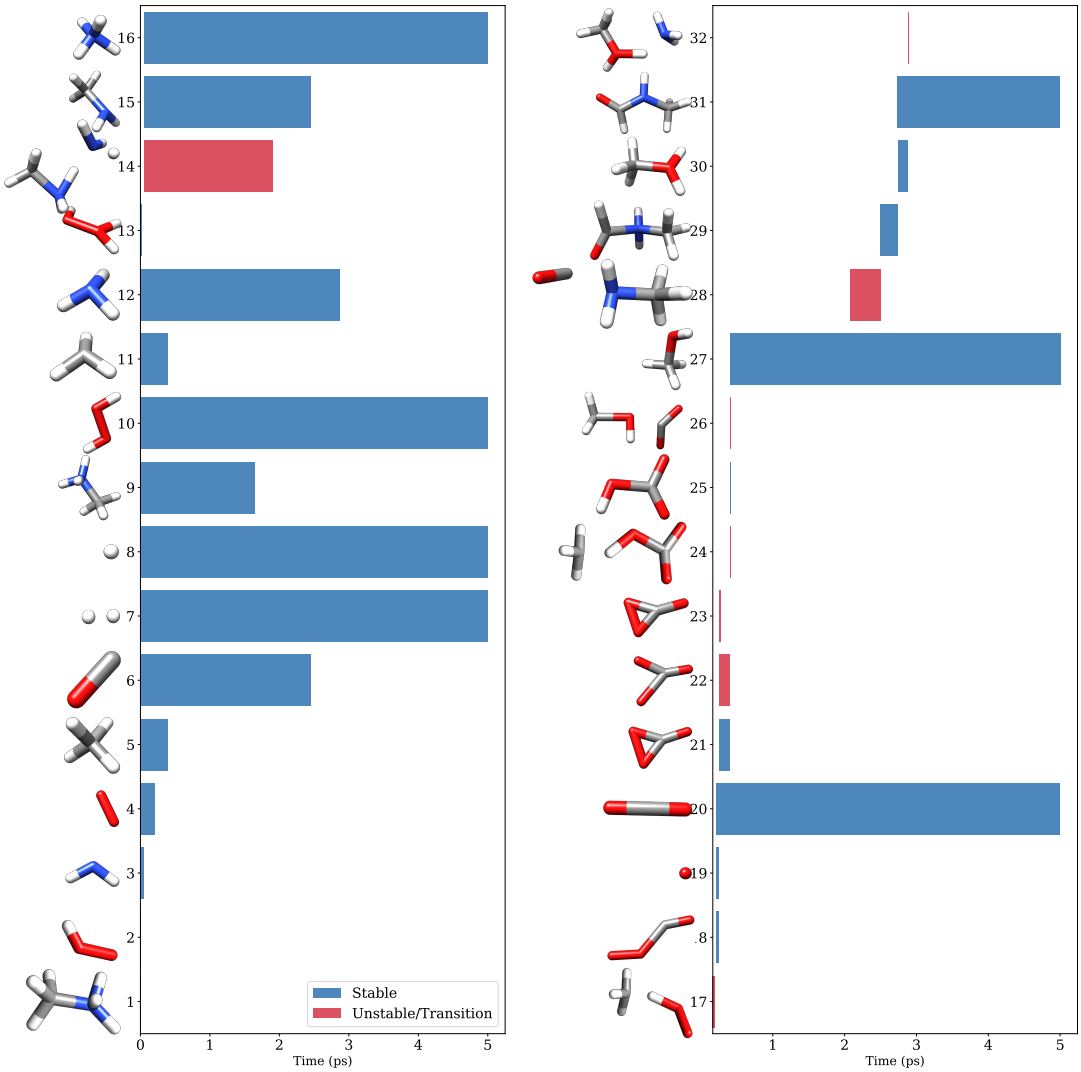


Figure 6.16: Lifetime analysis of molecular species observed during the nanoreactor simulation of the Urey–Miller experiment. Each bar represents a unique molecule identified by its MolBar identifier, with the length indicating the duration of its existence in the simulation.

Figure 6.16 and Figure 6.17 present the results from the nanoreactor simulation of the educational Urey–Miller experiment. The lifetime analysis shown in Figure 6.16 depicts the molecular species identified and their corresponding lifetimes, indicating the time of their initial formation and the duration of their existence before conversion into other species. This process required tracking the MolBar identifiers of each molecule at each time step to establish their first appearance and subsequent disappearance. The generation of InChI for the 32 unique molecules observed revealed that 18.75% (6 out of 32) of the molecules could not be assigned an InChI identifier. This is not due to limitations of InChI itself, but to errors in the conversion of 3D coordinates to Ctab files using `RDKit`[86] (version 2024.09.1), which were mainly caused by problems such as valence errors or atypical bond arrangements. This underscores the importance of incorporating 3D-to-bond connectivity conversion for identifiers used in electronic structure chemical space exploration, where chemically valid but unusual

structures frequently occur. It also demonstrates the benefit of employing non-valence-based algorithms, such as the graph neural network approach used by MolBar, for electronic hindrance analysis.

Under the accelerated simulation conditions, a wide range of species beyond the initial molecules formed within the first few steps. Early reactions often involved hydrogen abstractions or additions, yielding species such as O_2H (Molecule 2), NH_2 (Molecule 3), CH_3^+ (Molecule 11), NH_4^+ (Molecule 16), and OH (Molecule 7). H_2O disappeared immediately after the first step, transforming into hydrogen peroxide O_2H_2 (Molecule 12) and $O_2H_3^+$ (Molecule 13). Methane and ammonia quickly reacted to produce methylammonium CNH_5^+, first as an unstable intermediate with a highly distorted geometry (Molecule 1, where MolBar identified a CH bond from the ammonium hydrogen to the methane carbon, despite the unstable geometry), and later as a more stable molecule (Molecule 9) with a plausible geometry. A further example of H-abstraction is seen in structure 14, where ammonia interacts with methylammonium and removes a hydrogen, leading to the formation of methylamine CNH_4 (Molecule 15) and the ammonia ion NH_4^+ (Molecule 16). During the initial steps, carbon monoxide also reacted with dioxygen O_2 and atomic oxygen to create various species, such as carbon dioxide CO_2 (Molecule 20) and different CO_3 forms, including dioxiran-3-one (Molecule 21). Molecule 23 remained unstable and differed from Molecule 21 due to the non-planar configuration of its carbon atom, a feature detected by MolBar that became planar after optimization, thus changing the MolBar identifier.

The Urey–Miller experiment aimed to explore the chemical origins of life by replicating early Earth conditions and observing the synthesis of organic compounds, such as amino acids. In this simplified nanoreactor simulation, one amide, N-Methylformamide (C_2H_5NO), is formed (Molecule 31), which, while toxic, represents a simple amide and a potential intermediate in the formation of more complex amides, the building blocks of proteins. To examine the reaction pathway leading to N-Methylformamide, the reaction network in Figure 6.17 serves as a reference. This network is constructed by tracking the molecular affiliation of each atom at every time step using MolBar identifiers. A change in the MolBar for an atom is recorded as a reaction event, establishing connectivity between the molecules involved. The reaction network in Figure 6.17 depicts the relationships between molecular species observed during the simulation, where nodes correspond to unique molecules (identified by their MolBar identifiers) and edges represent observed reactions (changes in MolBar identifiers). To analyze the pathway leading to N-Methylformamide, the shortest path (see Section 2.3) from the initial molecules (red boxes) to N-Methylformamide (blue boxes) in the reaction network can be considered. The pathway is emphasized in Figure 6.17 with red edges, indicating the sequence of reactions leading to the formation of N-Methylformamide. The process begins with methylammonium (Molecule 1 and Molecule 9, derived from methane and ammonia). A hydrogen abstraction by another ammonia molecule yields methylamine (Molecule 15) and ammonium (Molecule 16). Subsequently, carbon monoxide (Molecule 6) coordinates to the nitrogen of methylamine via the carbon atom, forming an intermediate (Molecule 28). This intermediate

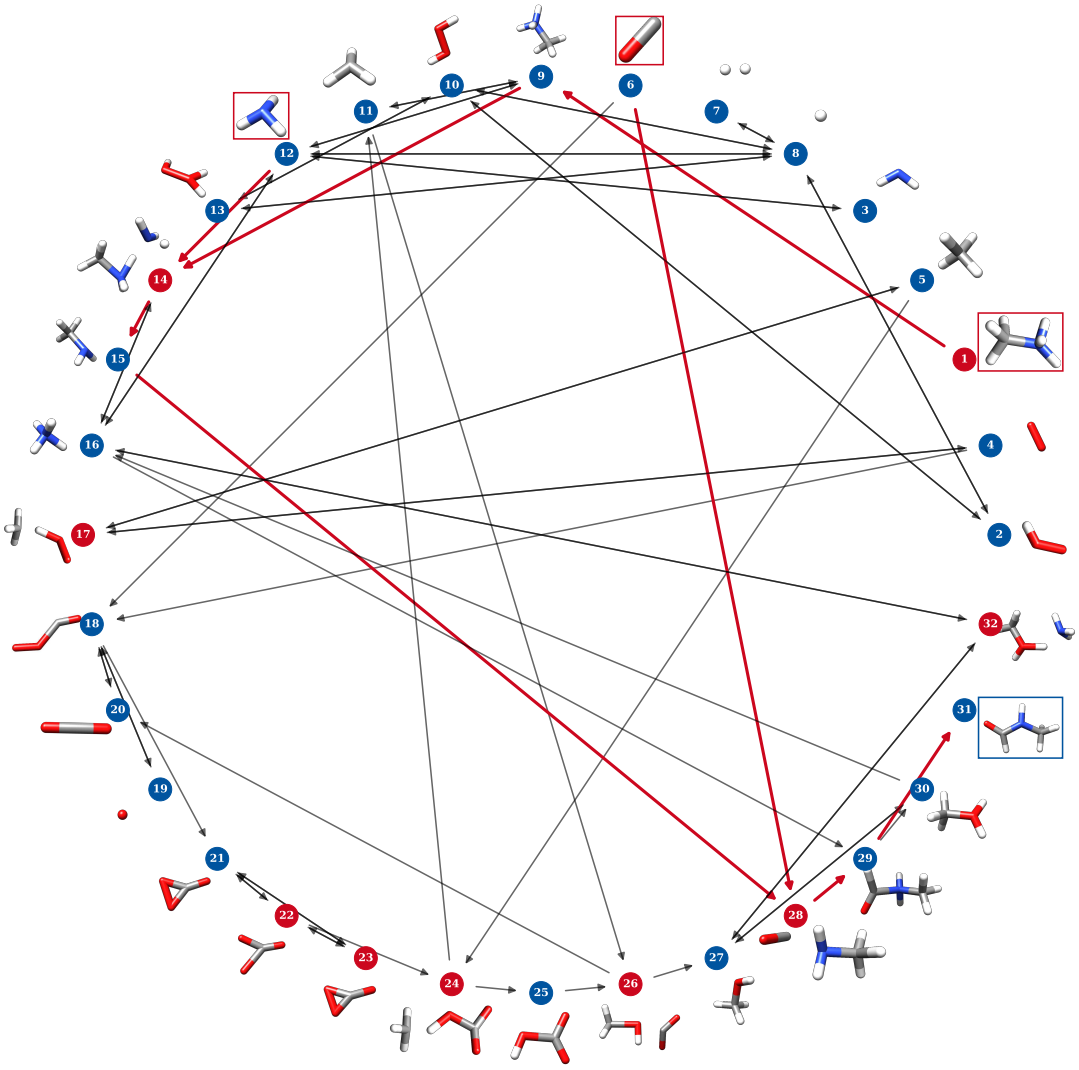


Figure 6.17: Reaction network generated from the nanoreactor simulation, illustrating the connectivity between different molecular species. Nodes represent unique molecules, while edges indicate observed reactions. The pathway leading to N-Methylformamide (Molecule 31) is highlighted with red edges, tracing the sequence of reactions from the initial molecules (red boxes) to the final product (blue box).

then produces protonated N-Methylformamide (Molecule 29). Deprotonation, facilitated by methanol (Molecule 27), results in the final formation of N-Methylformamide (Molecule 31).

All in all, MolBar represents the first identifier specifically designed for *ab initio* electronic structure chemical space exploration, naturally utilizing 3D coordinates as the primary input for analysis. This capability allows tools such as GOAT, CREST, and the *ab initio* nanoreactor to directly track molecular identities and transformations from the 3D structures generated during simulations, a feature particularly valuable in environments where chemically valid but unusual structures frequently emerge. Unlike widely used identifier algorithms such as SMILES and InChI, which depend on external tools to derive bond connectivity from 3D coordinates using valence rules, MolBar circumvents the limitations of this process, especially when unusual bonding patterns arise, a common occurrence in chemical space exploration. By offering a systematic approach to tracking molecular identities and transformations, MolBar accelerates the analysis of reaction discovery simulations, distinguishing between conformational changes, indicated by the same MolBar, and the formation of new molecules, indicated by a different MolBar, and enabling the construction of reaction networks based on observed transformations. Its ability to handle complex chemical structures and transformations enables chemical space exploration tools to go beyond ground-state organic chemistry and simple transition metal geometries and incorporate the chemistry of excited states, homogeneous organometallic and organocatalytic reactions, and systems with complex forms of chirality, including axial and planar chirality and atropisomerism.

7 Conclusion and Outlook

This work introduces a novel molecular identifier and line notation called the Molecular Barcode (MolBar). MolBar is purpose-built to integrate seamlessly with quantum chemistry and to interface directly with chemical space exploration tools that rely on 3D geometries and electronic structure theory. It is designed to extend beyond ground-state organic chemistry, standard organometallic chemistry, and atom-centered stereochemistry. To the best of the author's knowledge, MolBar offers broader coverage than existing molecular line notations for computational chemistry applications, particularly when compared to SMILES and InChI in this domain. Unlike standard cheminformatic identifiers, MolBar derives its information directly from the 3D geometry of the molecule, eliminating the need for external tools to determine connectivity, bond orders, or chirality. This ensures a rigorous one-to-one mapping between a specific 3D geometry and its MolBar string.

To capture stereochemical features that standard atom-based 2D graph representations cannot encode, MolBar conceptualizes the 3D shape of a molecule as a set of overlapping rigid fragments, each defined by its relative and absolute configuration. Structural information is then encoded using the eigenvalue spectra of three dedicated matrices: the topology matrix, the topography matrix, and the chirality matrix. The topology matrix represents molecular connectivity using a Hückel-like adjacency matrix. Since diastereomers differ in interatomic distances r_{ij}, the topography matrix captures molecular shape through a Coulomb-inspired matrix with $1/r_{ij}$ elements, enabling MolBar to describe highly complex stereoisomers, including those found in transition metal complexes. However, as distances also vary between conformers, a rigidity analysis is conducted to identify rigid substructures with minimal conformational flexibility. Rigidity is assessed based on ring structures, electronically hindered bonds with elevated bond orders (Wiberg bond order (WBO) ≥ 1.25), and sterically hindered single bonds exhibiting rotational barriers exceeding 20 kcal/mol. For electronically hindered bonds, a message passing neural network (MPNN) predicts WBO values; steric hindrance is identified using logistic regression trained on computational rotational barriers. Further small fluctuations in internal coordinates are filtered through a structure idealization step, which applies a specialized force field refinement to unify remaining matrix elements across conformers. Each rigid fragment is characterized by the Osipov-Pickup-Dunmur index, which enables the description of all stereoisomers including non-local chirality. The final MolBar string concatenates the eigenvalue spectra of these three matrices.

Across all benchmarks, MolBar matches or outperforms existing identifiers. The individual algorithms exhibit state-of-the-art or superior performance: the connectivity determination

algorithm was benchmarked against GFN-FF and `RDKit`'s CTD connectivity algorithms; the electronic hindrance algorithm was validated against WBO values derived from GFN2-xTB calculations; the steric hindrance algorithm was assessed using MLIP-based rotational barriers; and the force field geometry optimization was compared with standard optimizers from `CREST` and `SciPy`. The overall robustness of MolBar was further established in comparisons with InChI and SMILES across a range of tests, including single-example comparisons involving complex stereochemistry and deduplication tasks for large datasets. The permutation invariance of MolBar was confirmed using a dataset of 10,000 organic molecules. While the computational cost of MolBar exceeds that of InChI and SMILES, it remains lower than that of single-point calculations in electronic structure theory, making it well-suited for integration into chemical space exploration workflows that rely on such calculations. The practical utility of MolBar was demonstrated by analyzing MD trajectories and conformer ensembles to detect structural changes beyond mere conformational shifts, and by automatically constructing a reaction network during a nanoreactor simulation, where it successfully identified unique structures within the trajectory. Overall, this work represents a significant step toward standardized molecular identification in computational chemistry: by bridging the gap between cheminformatics conventions and electronic structure theory, MolBar provides a robust and comprehensive foundation for advanced chemical space exploration and quantum chemical databases.

Looking ahead, several research directions emerge naturally from the foundations established here. First, deriving MolBar directly from 2D representations would substantially broaden its utility, as most chemical databases store molecules as 2D structures. Currently, generating MolBar for these datasets requires computationally expensive and error-prone 2D-to-3D conversion. Since the rigidity analysis (rings, bond orders) does not inherently require 3D coordinates, small deflections per fragment combined with MolBar's internal force field could enable direct 2D derivation.

Second, identifying weak η-bonds in a robust manner remains an open challenge, particularly for transition metal complexes. Such bonds are often not recognized consistently across MD snapshots, leading to different MolBar strings for the same molecule. One solution is detecting weak η-bonds based on bond orders and proximity to metal centers, refining connectivity post hoc when distances exceed thresholds.

Third, enabling MolBar to be inverted back to a 3D structure would unlock new applications such as molecular structure storage and retrieval, as well as generative modeling. InChI and SMILES enable compression and reconstruction, making them suitable for molecular autoencoders and diffusion models. The MolBar matrices cannot be reconstructed from eigenvalues alone without eigenvectors; however, storing the flattened upper triangle of the sorted matrices as a string would enable full reconstruction of the 3D geometry, at the cost of a longer string. Compression techniques could reduce this overhead further.

Finally, extending MolBar to periodic systems—particularly polymers—represents a compelling

frontier. Existing extensions of SMILES and InChI are limited to simple polymers with a single repeating unit, failing to address complex polymers featuring multiple repeating units or branching structures. MolBar could be adapted to handle such complexity by characterizing the polymer's repeating unit and incorporating periodicity directly into the topology matrix, where the bond connecting the end of the repeating unit to its beginning is represented by a non-zero matrix element. Since this bond is typically a rotatable single bond, the topography and chirality matrices remain largely unaffected due to the fragmentation approach. The remaining challenge lies in representing the distribution of chain lengths and branching patterns, which could be addressed by treating mixtures of polymers as weighted combinations.

Bibliography

(1) G. N. Lewis, *J. Am. Chem. Soc.*, 1916, **38**, 762–785, DOI: 10.1007/s12045-019-0841-1.

(2) N. Bohr, *Philos. Mag.*, 1913, **26**, 1–25, DOI: 10.1016/b978-0-08-012103-1.50014-1.

(3) L. d. Broglie, *Philos. Mag.*, 1924, **47**, 446–458, DOI: 10.1080/09500830600914721.

(4) W. Heisenberg, *Z. Phys.*, 1925, **33**, 879–893, DOI: 10.1515/9783112596647-011.

(5) E. Schrödinger, *Ann. Phys.*, 1926, **385**, 437–490, DOI: 10.1515/9783112596647-008.

(6) W. Heisenberg, *Z. Phys.*, 1927, **43**, 172–198, DOI: 10.1007/978-3-642-61659-4_30.

(7) M. Born and J. R. Oppenheimer, *Ann. Phys.*, 1927, **389**, 457–484, DOI: 10.1002/andp.19273892002.

(8) W. Heitler and F. London, *Z. Phys.*, 1927, **44**, 455–472, DOI: 10.1007/bf01397394.

(9) W. Kutzelnigg, *Angew. Chem. Int. Ed.*, 1996, **35**, 572–586, DOI: 10.1002/anie.199605721.

(10) D. R. Hartree, Math. Proc. Cambridge Philos. Soc. 1928, vol. 24, pp. 89–110, DOI: 10.1017/s0305004100011919.

(11) J. C. Slater, *Phys. Rev.*, 1930, **35**, 210, DOI: 10.1103/physrev.35.210.2.

(12) V. Fock, *Z. Phys.*, 1930, **61**, 126–148, DOI: 10.1007/bf01340294.

(13) E. Hückel, *Z. Phys.*, 1931, **72**, 310–337, DOI: 10.1007/bf01341953.

(14) C. C. J. Roothaan, *Rev. Mod. Phys.*, 1951, **23**, 69, DOI: 10.1103/revmodphys.23.69.

(15) G. G. Hall, *Proc. R. Soc. Lond. A Math. Phys. Sci.*, 1951, **205**, 541–552, DOI: 10.1098/rspa.1951.0048.

(16) S. Boys, G. Cook, C. Reeves and I. Shavitt, *Nature*, 1956, **178**, 1207–1209, DOI: 10.1038/1781207a0.

(17) F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Hoboken, NJ, 2nd, 2017, ch. Introduction, pp. 1–2.

(18) R. Hoffmann, *J. Chem. Phys.*, 1963, **39**, 1397–1412, DOI: 10.1063/1.1734456.

(19) J. A. Pople and G. A. Segal, *J. Chem. Phys.*, 1965, **43**, S136–S151, DOI: 10.1063/1.1701476.

(20) J. A. Pople, D. Beveridge and P. Dobosh, *J. Chem. Phys.*, 1967, **47**, 2026–2033, DOI: 10.1063/1.1712233.

(21) M. J. Dewar and W. Thiel, *J. Am. Chem. Soc.*, 1977, **99**, 4899–4907, DOI: 10.1021/ja00457a004.

(22) P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864, DOI: 10.1103/physrev.136.b864.

(23) W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **137**, A1697, DOI: 10.1103/physrev.137.a1697.

(24) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian~16 Revision C.01*, Gaussian Inc. Wallingford CT, 2016.

(25) G. W. King, P. C. Cross and G. B. Thomas, *J. Chem. Phys.*, 1946, **14**, 35–42, DOI: 10.1063/1.1724059.

(26) T. Engel, in *Chemoinformatics: Basic Concepts and Methods*, ed. T. Engel and J. Gasteiger, Wiley-VCH, Weinheim, 1st, 2018, ch. 2, p. 4, DOI: 10.1002/9783527816880.ch2.

(27) P. D. Zemany, *Anal. Chem.*, 1950, **22**, 920–922, DOI: 10.1021/ac60043a021.

(28) L. Kuentzel, *Anal. Chem.*, 1951, **23**, 1413–1418, DOI: 10.1021/ac60058a016.

(29) W. L. Chen, *J. Chem. Inf. Model.*, 2006, **46**, 2230–2255, DOI: 10.1002/chin.200706208.

(30) D. Gluck, *J. Chem. Doc.*, 1965, **5**, 43–51, DOI: 10.1021/c160016a008.

(31) H. L. Morgan, *J. Chem. Doc.*, 1965, **5**, 107–113, DOI: 10.1021/c160017a018.

(32) E. H. Sussenguth, *J. Chem. Doc.*, 1965, **5**, 36–43, DOI: 10.1021/c160016a007.

(33) T. Engel and J. Gasteiger, *Chemoinformatics: Basic Concepts and Methods*, John Wiley & Sons, 2018, ch. Computer Processing of Chemical Structure Information, pp. 44–45.

(34) W. J. Wiswesser, *J. Chem. Inf. Comput. Sci.*, 1982, **22**, 88–93, DOI: 10.1021/ci00034a005.

(35) D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36, DOI: 10.1021/ci00057a005.

(36) D. Weininger, A. Weininger and J. L. Weininger, *J. Chem. Inf. Comput. Sci.*, 1989, **29**, 97–101, DOI: 10.1021/ci00062a008.

(37) S. Ash, M. A. Cline, R. W. Homer, T. Hurst and G. B. Smith, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 71–79, DOI: 10.1021/ci960109j.

(38) J. M. Barnard, C. J. Jochum and S. M. Welford, in *Chemical Structure Information Systems*, ed. W. A. Warr, ACS Publications, 1989, ch. 8, pp. 76–81, DOI: 10.1021/bk-1989-0400.ch008.

(39) S. R. Heller, A. McNaught, I. Pletnev, S. Stein and D. Tchekhovskoi, *J. Cheminformatics*, 2015, **7**, 1–34, DOI: 10.1186/s13321-015-0068-4.

(40) F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 73–78.

(41) F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka and F. Weigend, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2014, **4**, 91–100, DOI: 10.1002/wcms.1162.

(42) S. Seritan, C. Bannwarth, B. S. Fales, E. G. Hohenstein, C. M. Isborn, S. I. Kokkila-Schumacher, X. Li, F. Liu, N. Luehr, J. W. Snyder Jr et al., *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2021, **11**, e1494, DOI: 10.1002/wcms.1494.

(43) A. I. Krylov and P. M. Gill, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2013, **3**, 317–326, DOI: 10.1002/wcms.1122.

(44) P. Pracht, E. Caldeweyher, S. Ehlert and S. Grimme, *ChemRxiv*, 2019, DOI: 10.26434/chemrxiv.8326202.v1.

(45) S. Grimme, C. Bannwarth and P. Shushkov, *J. Chem. Theory Comput.*, 2017, **13**, 1989–2009, DOI: 10.1021/acs.jctc.7b00118.

(46) C. Bannwarth, S. Ehlert and S. Grimme, *J. Chem. Theory Comput.*, 2019, **15**, 1652–1671, DOI: 10.1021/acs.jctc.8b01176.

(47) B. de Souza, *Angew. Chem. Int. Ed.*, 2025, **64**, e202500393, DOI: 10.1002/ange.202500393.

(48) P. Pracht, F. Bohle and S. Grimme, *Phys. Chem. Chem. Phys.*, 2020, **22**, 7169–7192, DOI: 10.1039/c9cp06869d.

(49) C. Zurek, R. A. Mallaev, A. C. Paul, N. van Staalduinen, P. Pracht, R. Ellerbrock and C. Bannwarth, *J. Chem. Theory Comput.*, 2025, **21**, 1459–1475, DOI: 10.1021/acs.jctc.4c01275.

(50) L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V. S. Pande and T. J. Martínez, *Nat. Chem.*, 2014, **6**, 1044–1048, DOI: 10.1038/nchem.2099.

(51) J. P. Unsleber, S. A. Grimmel and M. Reiher, *J. Chem. Theory Comput.*, 2022, **18**, 5393–5409, DOI: 10.1021/acs.jctc.2c00193.

(52) J. Gorges and S. Grimme, *Phys. Chem. Chem. Phys.*, 2025, **27**, 6899–6911, DOI: 10.1039/D5CP00316D.

(53) Y. Chen and C. Bannwarth, *J. Chem. Inf. Model.*, 2025, **65**, 9125–9141, DOI: 10.1021/acs.jcim.5c00908.

(54) M. Nakata and T. Shimazaki, *J. Chem. Inf. Model.*, 2017, **57**, 1300–1308, DOI: 10.1021/acs.jcim.7b00083.

(55) R. Ramakrishnan, P. O. Dral, M. Rupp and O. A. Von Lilienfeld, *Sci. Data*, 2014, **1**, 1–7, DOI: 10.1038/sdata.2014.22.

(56) C. Isert, K. Atz, J. Jiménez-Luna and G. Schneider, *Sci. Data*, 2022, **9**, 273, DOI: 10.1038/s41597-022-01390-7.

(57) D. Balcells and B. B. Skjelstad, *J. Chem. Inf. Model.*, 2020, **60**, 6135–6146, DOI: 10.1021/acs.jcim.0c01041.

(58) D. S. Levine, M. Shuaibi, E. W. C. Spotte-Smith, M. G. Taylor, M. R. Hasyim, K. Michel, I. Batatia, G. Csányi, M. Dzamba, P. Eastman, N. C. Frey, X. Fu, V. Gharakhanyan, A. S. Krishnapriyan, J. A. Rackers, S. Raja, A. Rizvi, A. S. Rosen, Z. Ulissi, S. Vargas, C. L. Zitnick, S. M. Blau and B. M. Wood, *arXiv*, 2025, DOI: 10.48550/arXiv.2505.08762.

(59) G. A. Pinheiro, J. Mucelini, M. D. Soares, R. C. Prati, J. L. Da Silva and M. G. Quiles, *J. Phys. Chem. A*, 2020, **124**, 9854–9866, DOI: 10.1021/acs.jpca.0c05969.

(60) A. G. Garrison, J. Heras-Domingo, J. R. Kitchin, G. dos Passos Gomes, Z. W. Ulissi and S. M. Blau, *J. Chem. Inf. Model.*, 2023, **63**, 7642–7654, DOI: 10.1021/acs.jcim.3c01226.

(61) J. S. Smith, O. Isayev and A. E. Roitberg, *Chem. Sci.*, 2017, **8**, 3192–3203, DOI: 10.1039/c6sc05720a.

(62) X. Gao, F. Ramezanghorbani, O. Isayev, J. S. Smith and A. E. Roitberg, *J. Chem. Inf. Model.*, 2020, **60**, 3408–3415, DOI: 10.1021/acs.jcim.0c00451.

(63) K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko and K.-R. Müller, *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*, 2017, DOI: 10.48550/arXiv.1706.08566.

(64) I. Batatia, D. P. Kovács, G. N. C. Simm, C. Ortner and G. Csányi, *MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields*, 2023, DOI: 10.48550/arXiv.2206.07697.

(65) B. M. Wood, M. Dzamba, X. Fu, M. Gao, M. Shuaibi, L. Barroso-Luque, K. Abdelmaqsoud, V. Gharakhanyan, J. R. Kitchin, D. S. Levine et al., *arXiv*, 2025, DOI: 10.48550/arXiv.2506.23971.

(66) F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Hoboken, NJ, 2nd, 2017, ch. The Dynamical Equation, pp. 7–8.

(67) F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Hoboken, NJ, 2nd, 2017, ch. Configuration Interaction, pp. 137–147.

(68) F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Hoboken, NJ, 2nd, 2017, ch. Density Functional Theory, pp. 235–255.

(69) T. Froitzheim, M. Müller, A. Hansen and S. Grimme, *ChemRxiv*, 2025, DOI: 10.26434/chemrxiv-2025-bjxvt.

(70) C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher and S. Grimme, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2021, **11**, e1493, DOI: 10.1002/wcms.1493.

(71) F. Jensen, *Introduction to Computational Chemistry*, John Wiley & Sons, Hoboken, NJ, 2nd, 2017, ch. Vibrational Degrees of Freedom, pp. 432–433.

(72) E. Lewars, *Computational chemistry: introduction to the theory and applications of molecular and quantum mechanics*, Springer, New York, NY, 3rd, 2016, ch. The Concept of the Potential Energy Surface, pp. 9–22, DOI: 10.1007/978-3-031-51443-2.

(73) H. B. Schlegel, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 790–809, DOI: 10.1002/wcms.34.

(74) M. Ōki, *Top. Stereochem.*, 1983, 1–81, DOI: 10.1002/9780470147238.ch1.

(75) Matgenix, *The coord file*, Accessed: 2026-01-07, 2026.

(76) T. Engel and J. Gasteiger, *Chemoinformatics: Basic Concepts and Methods*, John Wiley & Sons, 2018, ch. Computer Processing of Chemical Structure Information, pp. 50–51.

(77) A. Dalby, J. G. Nourse, W. D. Hounshell, A. K. Gushurst, D. L. Grier, B. A. Leland and J. Laufer, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 244–255, DOI: 10.1021/ci00007a012.

(78) D. S. BIOVIA, *CT File Formats*, Accessed: 2026-01-07, 2020.

(79) H. M. Berman, G. J. Kleywegt, H. Nakamura and J. L. Markley, *J. Comput. Aided Mol. Des.*, 2014, **28**, 1009–1014, DOI: 10.1007/s10822-014-9770-y.

(80) I. D. Brown and B. McMahon, *Struct. Sci.*, 2002, **58**, 317–324, DOI: 10.1107/s0108768102003464.

(81) P. Murray-Rust and H. S. Rzepa, *J. Cheminformatics*, 2011, **3**, 44, DOI: 10.1186/1758-2946-3-44.

(82) I. Daylight Chemical Information Systems, *SMILES - A Simplified Chemical Language*, Accessed: 2026-01-16, 2026.

(83) S. Heller, A. McNaught, S. Stein, D. Tchekhovskoi and I. Pletnev, *J. Cheminformatics*, 2013, **5**, 7, DOI: 10.1186/1758-2946-5-7.

(84) S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu et al., *Nucleic Acids Res.*, 2023, **51**, D1373–D1380, DOI: 10.1093/nar/gkae1059.

(85) N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Cheminformatics*, 2011, **3**, 33, DOI: 10.1186/1758-2946-3-33.

(86) RDKit Community, *RDKit: Open-source cheminformatics*, Accessed: 25 September 2024, 2024.

(87) G. Blanke, J. Brammer, D. Baljozovic, N. U. Khan, F. Lange, F. Bänsch, C. A. Tovee, U. Schatzschneider, R. M. Hartshorn and S. Herres-Pawlis, *Faraday Discuss.*, 2025, **256**, 503–519, DOI: 10.1039/d4fd00145a.

(88) W. J. Wiswesser, *Chem. Eng. News*, 1952, **30**, 3523–3526, DOI: 10.1021/cen-v030n034.p3523.

(89) W. J. Wiswesser, *J. Chem. Doc.*, 1968, **8**, 146–150, DOI: 10.1021/c160030a007.

(90) R. W. Homer, J. Swanson, R. J. Jilek, T. Hurst and R. D. Clark, *J. Chem. Inf. Model.*, 2008, **48**, 2294–2307, DOI: 10.1021/ci7004687.

(91) M. Krenn, F. Häse, A. Nigam, P. Friederich and A. Aspuru-Guzik, *Mach. Learn.: Sci. Technol.*, 2020, **1**, 045024, DOI: 10.1088/2632-2153/aba947.

(92) A. H. Cheng, A. Cai, S. Miret, G. Malkomes, M. Phielipp and A. Aspuru-Guzik, *Digit. Discov.*, 2023, **2**, 748–758, DOI: 10.1039/D3DD00012E.

(93) J.-G. Sobez and M. Reiher, *J. Chem. Inf. Model.*, 2020, **60**, 3884–3900, DOI: 10.1021/acs.jcim.0c00503.

(94) T. Engel and J. Gasteiger, *Chemoinformatics: Basic Concepts and Methods*, John Wiley & Sons, 2018, ch. Principles of Molecular Representations, p. 15.

(95) E. W. Dijkstra, *Numer. Math.*, 1959, **1**, 269–271, DOI: 10.1145/3544585.3544600.

(96) R. W. Floyd, *Commun. ACM*, 1962, **5**, 345–345, DOI: 10.1145/367766.368168.

(97) S. Warshall, *J. ACM*, 1962, **9**, 11–12, DOI: 10.1145/321105.321107.

(98) T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to algorithms*, MIT Press, Cambridge, MA, 3rd, 2009, ch. 24, pp. 658–664, DOI: 10.1007/978-1-4842-7077-6_3.

(99) M. L. Fredman and R. E. Tarjan, *J. ACM*, 1987, **34**, 596–615, DOI: 10.1145/28869.28874.

(100) A. Schrijver, *Doc. Math.*, 2012, **17**, 155–167, DOI: 10.4171/dms/6/19.

(101) T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to algorithms*, MIT Press, Cambridge, MA, 3rd, 2009, ch. 25, pp. 693–699, DOI: 10.1007/978-1-4842-7077-6_3.

(102) T. H. Cormen, C. E. Leiserson, R. L. Rivest and C. Stein, *Introduction to algorithms*, MIT Press, Cambridge, MA, 3rd, 2009, ch. 22, pp. 594–602, DOI: 10.1007/978-1-4842-7077-6_3.

(103) B. T. Fan, A. Panaye, J. P. Doucet and A. Barbu, *J. Chem. Inf. Comput. Sci.*, 1993, **33**, 657–662, DOI: 10.1021/ci00015a002.

(104) J. Figueras, *J. Chem. Inf. Comput. Sci.*, 1996, **36**, 986–991, DOI: 10.1021/ci960013p.

(105) T. Hanser, P. Jauffret and G. Kaufmann, *J. Chem. Inf. Comput. Sci.*, 1996, **36**, 1146–1152, DOI: 10.1021/ci960322f.

(106) S. Grimme, *Angew. Chem. Int. Ed.*, 2013, **52**, 6306–6312, DOI: 10.1002/anie.201300158.

(107) C. A. Bauer and S. Grimme, *J. Phys. Chem. A*, 2016, **120**, 3755–3766, DOI: 10.1021/acs.jpca.6b02907.

(108) V. Ásgeirsson, B. O. Birgisson, R. Bjornsson, U. Becker, F. Neese, C. Riplinger and H. Jónsson, *J. Chem. Theory Comput.*, 2021, **17**, 4929–4945, DOI: 10.1021/acs.jctc.1c00462.

(109) P. Zimmerman, *J. Chem. Theory Comput.*, 2013, **9**, 3043–3050, DOI: 10.1021/ct400319w.

(110) E. Martínez-Núñez, *Phys. Chem. Chem. Phys.*, 2015, **17**, 14912–14921, DOI: 10.1039/c5cp02175h.

(111) A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*, Courier Corporation, 2012, pp. 151–151.

(112) László Poppe and József Nagy and Gábor Hornyánszky and Zoltán Boros, *Stereochemistry and Stereoselective Synthesis: An Introduction*, Wiley-VCH Verlag, 2016, ch. ConceptsofStereochemistry, pp. 26–27, DOI: 10.1002/9783527341177.

(113) S. R. LaPlante, P. J. Edwards, L. D. Fader, A. Jakalian and O. Hucke, *ChemMedChem*, 2011, **6**, 505–513, DOI: 10.1002/cmdc.201000485.

(114) S. J. Blanksby and G. B. Ellison, *Acc. Chem. Res.*, 2003, **36**, 255–263, DOI: 10.1002/chin.200324299.

(115) E. Iglesias, *J. Org. Chem.*, 2003, **68**, 2680–2688, DOI: 10.1021/jo026629x.

(116) W. Tang, T. Yang, C. A. Morales-Rivera, X. Geng, V. K. Srirambhatla, X. Kang, V. P. Chauhan, S. Hong, Q. Tu, A. J. Florence et al., *Nat. Commun.*, 2023, **14**, 561, DOI: 10.1038/s41467-023-35924-3.

(117) Y. Chiang, A. Kresge and N. Schepp, *J. Am. Chem. Soc.*, 1989, **111**, 3977–3980, DOI: 10.1021/ja00193a033.

(118) J. S. Al-Otaibi, *SpringerPlus*, 2015, **4**, 586, DOI: 10.1186/s40064-015-1363-2.

(119) F. Leroux, *ChemBioChem*, 2004, **5**, 644–649, DOI: 10.1002/cbic.200300906.

(120) K. B. Wiberg and P. R. Rablen, *J. Am. Chem. Soc.*, 1995, **117**, 2201–2209, DOI: 10.1021/ja00113a009.

(121) M. Quirós, S. Gražulis, S. Girdzijauskaitė, A. Merkys and A. Vaitkus, *J. Cheminformatics*, 2018, **10**, 1–17, DOI: 10.1186/s13321-018-0279-6.

(122) M. H. Rasmussen, M. Strandgaard, J. Seumer, L. K. Hemmingsen, A. Frei, D. Balcells and J. H. Jensen, *J. Cheminform.*, 2025, **17**, 1–13, DOI: 10.1186/s13321-025-01008-1.

(123) M. Iwamura, S. Takeuchi and T. Tahara, *J. Am. Chem. Soc.*, 2007, **129**, 5248–5256, DOI: 10.1021/ja069300s.

(124) C. E. Housecroft and E. C. Constable, *J. Mater. Chem. C*, 2022, **10**, 4456–4482, DOI: 10.1039/D1TC04028F.

(125) László Poppe and József Nagy and Gábor Hornyánszky and Zoltán Boros, in *Stereochemistry and Stereoselective Synthesis: An Introduction*, ed. László Poppe and József Nagy, Wiley-VCH, Weinheim, 2018, ch. 2, pp. 44–46, DOI: 10.1002/9783527341177.

(126) R. Merris, *Linear Algebra Appl.*, 1994, **197**, 143–176, DOI: 10.1016/0024-3795(94)90486-3.

(127) G. Strang, *Linear Algebra and Its Applications*, Harcourt Brace Jovanovich, San Diego, 1988.

(128) K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko and K.-R. Müller, *J. Chem. Theory Comput.*, 2013, **9**, 3404–3419, DOI: 10.1021/ct400195d.

(129) G. Ferré, T. Haut and K. Barros, *J. Chem. Phys.*, 2017, **146**, 114107, DOI: 10.1063/1.4978623.

(130) F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley and O. A. von Lilienfeld, *J. Chem. Theory Comput.*, 2017, **13**, 5255–5264, DOI: 10.1021/acs.jctc.7b00577.

(131) M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. Von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301, DOI: 10.1103/physrevlett.108.058301.

(132) J. Schrier, *J. Chem. Inf. Model.*, 2020, **60**, 3804–3811, DOI: 10.1021/acs.jcim.0c00631.

(133) A. A. Hagberg, D. A. Schult and P. J. Swart, Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, 2008, pp. 11–15, DOI: 10.25080/tcwv9851.

(134) C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke and T. E. Oliphant, *Nature*, 2020, **585**, 357–362, DOI: 10.1038/s41586-020-2649-2.

(135) P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt and SciPy 1.0 Contributors, *Nat. Methods*, 2020, **17**, 261–272, DOI: 10.1038/s41592-019-0686-2.

(136) Jensen Group, *xyz2mol*, Accessed: 25 September 2024, 2024.

(137) K. Alhameedi, G. S. Chandler and D. Jayatilaka, *Results Chem.*, 2020, **2**, 100053, DOI: 10.1016/j.rechem.2020.100053.

(138) R. S. Mulliken, *J. Chem. Phys.*, 1955, **23**, 1833–1840, DOI: 10.1063/1.1740588.

(139) K. B. Wiberg, *Tetrahedron*, 1968, **24**, 1083–1096, DOI: 10.1016/0040-4020(68)88057-3.

(140) I. Mayer, *Chem. Phys. Lett.*, 1983, **97**, 270–274, DOI: 10.1016/0009-2614(85)85253-2.

(141) R. Bader, W. H. Henneker and P. E. Cade, *J. Chem. Phys.*, 1967, **46**, 3341–3363, DOI: 10.1063/1.1841222.

(142) A. D. Becke and K. E. Edgecombe, *J. Chem. Phys.*, 1990, **92**, 5397–5403, DOI: 10.1063/1.458517.

(143) A. Lüchow, *J. Comput. Chem.*, 2014, **35**, 854–864, DOI: 10.1002/jcc.23561.

(144) B. Cordero, V. Gómez, A. E. Platero-Prats, M. Revés, J. Echeverría, E. Cremades, F. Barragán and S. Alvarez, *Dalton T.*, 2008, **37**, 2832–2838, DOI: 10.1039/B801115J.

(145) G. A. Landrum and W. V. Glassey, *YAeHMOP: Yet Another extended Hückel Molecular Orbital Package*, http://yaehmop.sourceforge.net/, Accessed: 2025-10-24, 2025.

(146) S. Spicher and S. Grimme, *Angew. Chem. Int. Ed.*, 2020, **59**, 15665–15673, DOI: 10.1002/ange.202004239.

(147) S. Grimme, J. G. Brandenburg, C. Bannwarth and A. Hansen, *J. Chem. Phys.*, 2015, **143**, 054107, DOI: 10.1063/1.4927476.

(148) J. G. Brandenburg, C. Bannwarth, A. Hansen and S. Grimme, *J. Chem. Phys.*, 2018, **148**, 064104, DOI: 10.1063/1.5012601.

(149) S. Spicher, M. Bursch and S. Grimme, *J. Phys. Chem. C*, 2020, **124**, 27529–27541, DOI: 10.1021/acs.jpcc.0c08617.

(150) M. Bursch, A. Hansen, P. Pracht, J. T. Kohn and S. Grimme, *Phys. Chem. Chem. Phys.*, 2021, **23**, 287–299, DOI: 10.1039/d0cp04696e.

(151) S. Spicher and S. Grimme, *J. Phys. Chem. Lett.*, 2020, **11**, 6606–6611, DOI: 10.1021/acs.jpclett.0c01930.

(152) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104, DOI: 10.1063/1.3382344.

(153) P. Pyykkö and M. Atsumi, *Chem. Eur. J.*, 2009, **15**, 186–197, DOI: 10.1002/chem.200800987.

(154) N. V. Staalduinen and C. Bannwarth, *J. Chem. Inf. Model.*, 2026, **66**, 10935–10955, DOI: 10.1021/acs.jcim.6c00688.

(155) R. A. Olsen, L. Liu, N. Ghaderi, A. Johns, M. E. Hatcher and L. J. Mueller, *J. Am. Chem. Soc.*, 2003, **125**, 10125–10132, DOI: 10.1021/ja028751j.

(156) K. E. Laidig and L. M. Cameron, *J. Am. Chem. Soc.*, 1996, **118**, 1737–1742, DOI: 10.1021/ja952678y.

(157) H. Kneiding, R. Lukin, L. Lang, S. Reine, T. B. Pedersen, R. De Bin and D. Balcells, *Digit. Discov.*, 2023, **2**, 618–633, DOI: 10.1039/d2dd00129b.

(158) J. C. Baber and E. E. Hodgkin, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 401–406, DOI: 10.1021/ci00009a001.

(159) M. Hendlich, F. Rippmann and G. Barnickel, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 774–778, DOI: 10.1021/ci9603487.

(160) R. Sayle, *PDB: Cruft to Content*, Accessed: 2025-11-13, 2001.

(161) P. Labute, *J. Chem. Inf. Model.*, 2005, **45**, 215–221, DOI: 10.1021/ci049915d.

(162) Y. Kim and W. Y. Kim, *Bull. Korean Chem. Soc.*, 2015, **36**, 1769–1777, DOI: 10.1002/bkcs.10334.

(163) T. A. Soares, A. Nunes-Alves, A. Mazzolari, F. Ruggiu, G.-W. Wei and K. Merz, *The (Re)-Evolution of Quantitative Structure–Activity Relationship (QSAR) studies propelled by the surge of machine learning methods*, 2022, DOI: 10.1021/acs.jcim.2c01422.

(164) J. P. Janet and H. J. Kulik, in *Machine Learning in Chemistry*, ed. J. P. Janet and H. J. Kulik, American Chemical Society, Washington, DC, 1st, 2020, ch. 2, pp. 44–47.

(165) Q. Wang, Y. Ma, K. Zhao and Y. Tian, *Ann. Data Sci.*, 2022, **9**, 187–212, DOI: 10.1007/s40745-020-00253-5.

(166) T. Hastie, R. Tibshirani and J. Friedman, in *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, ed. T. Hastie, R. Tibshirani and J. Friedman, Springer, New York, NY, 2nd, 2018, ch. 3, p. 44.

(167) C. Williams and C. Rasmussen, *Advances in neural information processing systems*, 1995, **8**, 514–520.

(168) L. Breiman, *Mach. Learn.*, 2001, **45**, 5–32, DOI: 10.1023/a:1010933404324.

(169) Y. LeCun, Y. Bengio and G. Hinton, *Nature*, 2015, **521**, 436–444, DOI: 10.1038/nature14539.

(170) D. E. Rumelhart, G. E. Hinton and R. J. Williams, *Nature*, 1986, **323**, 533–536, DOI: 10.7551/mitpress/1888.003.0013.

(171) N. Ketkar and J. Moolayil, in *Deep Learning with python: learn best practices of deep learning models with PyTorch*, Springer, 2021, pp. 133–145, DOI: 10.1007/978-1-4842-5364-9_4.

(172) S. Ioffe and C. Szegedy, *arXiv*, 2015, DOI: 10.48550/arXiv.1502.03167.

(173) J. L. Ba, J. R. Kiros and G. E. Hinton, *arXiv*, 2016, DOI: 10.48550/arXiv.1607.06450.

(174) S. Sharma, S. Sharma and A. Athaiya, *IJEAST*, 2017, **6**, 310–316, DOI: 10.33564/ijeast.2020.v04i12.054.

(175) P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer et al., *Commun. Mater.*, 2022, **3**, 93, DOI: 10.1038/s43246-022-00315-6.

(176) J. Gasteiger, J. Groß and S. Günnemann, *arXiv*, 2020, DOI: 10.48550/arXiv.2003.03123.

(177) J. Gasteiger, C. Yeshwanth and S. Günnemann, *Directional Message Passing on Molecular Graphs via Synthetic Coordinates*, 2022, DOI: 10.48550/arXiv.2111.04718.

(178) O. T. Unke and M. Meuwly, *J. Chem. Theory Comput.*, 2019, **15**, 3678–3693, DOI: 10.1021/acs.jctc.9b00181.

(179) N. Lubbers, J. S. Smith and K. Barros, *arXiv*, 2020, DOI: 10.48550/arXiv.1710.00017.

(180) K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea et al., *J. Chem. Inf. Model.*, 2019, **59**, 3370–3388, DOI: 10.1021/acs.jcim.9b00237.

(181) E. Heid, K. P. Greenman, Y. Chung, S.-C. Li, D. E. Graff, F. H. Vermeire, H. Wu, W. H. Green and C. J. McGill, *J. Chem. Inf. Model.*, 2023, **64**, 9–17, DOI: 10.1021/acs.jcim.3c01250.

(182) V. G. Satorras, E. Hoogeboom and M. Welling, *arXiv*, 2022, DOI: 10.48550/arXiv.2102.09844.

(183) J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, *Neural Message Passing for Quantum Chemistry*, 2017.

(184) N. Lubbers, J. S. Smith and K. Barros, *J. Chem. Phys.*, 2018, **148**, 241717, DOI: 10.1063/1.5011181.

(185) S.-C. Li, H. Wu, A. Menon, K. A. Spiekermann, Y.-P. Li and W. H. Green, *J. Am. Chem. Soc.*, 2024, **146**, 23103–23120, DOI: 10.1021/jacs.4c04670.

(186) S. Magedov, C. Koh, W. Malone, N. Lubbers and B. Nebgen, *J. Appl. Phys.*, 2021, **129**, 064701, DOI: 10.1063/5.0016011.

(187) W. Kutzelnigg, *Angew. Chem. Int. Ed.*, 1984, **23**, 272–295, DOI: 10.1002/anie.198402721.

(188) P. P. Power, *Organometallics*, 2020, **39**, 4127–4138, DOI: 10.1021/acs.organomet.0c00200.

(189) N. van Staalduinen and C. Bannwarth, *Digit. Discov.*, 2024, **3**, 2298–2319, DOI: 10.1039/d4dd00208c.

(190) A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, 2019, DOI: 10.48550/arXiv.1912.01703.

(191) M. Fey and J. E. Lenssen, *arXiv*, 2019, DOI: 10.48550/arXiv.1903.02428.

(192) Pat Walters, *Some Thoughts on Splitting Chemical Datasets*, Accessed: 24 January 2026, 2026.

(193) G. W. Bemis and M. A. Murcko, *J. Med. Chem.*, 1996, **39**, 2887–2893, DOI: 10.1021/jm9602928.

(194) Q. Guo, S. Hernandez-Hernandez and P. J. Ballester, International Conference on Artificial Neural Networks, 2024, pp. 58–72, DOI: 10.1007/978-3-031-72359-9_5.

(195) L. McInnes, J. Healy and J. Melville, *arXiv*, 2018, DOI: 10.48550/arXiv.1802.03426.

(196) J. group, *xyz2mol_tm*, https://github.com/jensengroup/xyz2mol_tm/, Accessed: 2025-11-24, 2025.

(197) M. S. Betson, J. Clayden, C. P. Worrall and S. Peace, *Angew. Chem. Int. Ed.*, 2006, **45**, 5803–5807, DOI: 10.1002/anie.200601866.

(198) M. Basilaia, M. H. Chen, J. Secka and J. L. Gustafson, *Acc. Chem. Res.*, 2022, **55**, 2904–2919, DOI: 10.1021/acs.accounts.2c00500.

(199) G. Christie and J. Kenner, *J. Chem. Soc., Trans*, 1922, **121**, 614, DOI: 10.1039/JR9262900671.

(200) K. Mislow, *Angew. Chem.*, 1958, **70**, 683–689, DOI: 10.1002/ange.19580702203.

(201) S. R. LaPlante, L. D. Fader, K. R. Fandrick, D. R. Fandrick, O. Hucke, R. Kemper, S. P. Miller and P. J. Edwards, *J. Med. Chem.*, 2011, **54**, 7005–7022, DOI: 10.1021/jm200584g.

(202) M. Melchert and A. List, *Int. J. Biochem. Cell Biol.*, 2007, **39**, 1489–1499, DOI: 10.1016/j.biocel.2007.01.022.

(203) B. A. Lanman, A. T. Parsons and S. G. Zech, *Acc. Chem. Res.*, 2022, **55**, 2892–2903, DOI: 10.1021/acs.accounts.2c00479.

(204) C. Yang, D. Zhou, Z. Shen, D. M. Wilson, M. Renner, J. N. Miner, J.-L. Girardet and C. A. Lee, *Drug Metab. Dispos.*, 2019, **47**, 104–113, DOI: 10.1124/dmd.118.080549.

(205) D. S. BIOVIA, *Chemical Representation 2025*, Accessed: 2026-02-04, 2025.

(206) C. D. Discovery, *Advanced Stereochemistry Registration: Atropisomers, Mixtures, Unknowns and Non-Tetrahedral Chirality*, Accessed: 2026-02-04, 2026.

(207) R. M. Hanson, *J. Cheminformatics*, 2016, **8**, 1–20, DOI: 10.1186/s13321-016-0160-4.

(208) ChemAxon, *ChemAxon Extended SMILES and SMARTS (CXSMILES and CXSMARTS)*, Accessed: 2025-06-30, 2024.

(209) J. Clayden and J. H. Pink, *Angew. Chem. Int. Ed.*, 1998, **37**, 1937–1939, DOI: 10.1002/(sici)1521-3773(19980803)37:13/14<1937::aid-anie1937>3.0.co;2-4.

(210) T. Balduf, P. A. Gerken, M. Y. Shelley, M. A. Watson, M. C. Bennett, M. Svensson, A. E. Leffler and A. Bochevarov, *J. Chem. Inf. Model.*, 2026, DOI: 10.1021/acs.jcim.5c02720.

(211) G. Bott, L. D. Field and S. Sternhell, *J. Am. Chem. Soc.*, 1980, **102**, 5618–5626, DOI: 10.1021/ja00537a036.

(212) V. Belot, D. Farran, M. Jean, M. Albalat, N. Vanthuyne and C. Roussel, *J. Org. Chem.*, 2017, **82**, 10188–10200, DOI: 10.1021/acs.joc.7b01698.

(213) M. N. Khan and T. Wirth, *Chem. Eur. J.*, 2021, **27**, 7059–7068, DOI: 10.1002/chem.202005317.

(214) M. Flos, P. Lameiras, C. Denhez, C. Mirand and H. Berber, *J. Org. Chem.*, 2016, **81**, 2372–2382, DOI: 10.1021/acs.joc.5b02856.

(215) A. Ciogli, S. V. Kumar, M. Mancinelli, A. Mazzanti, S. Perumal, C. Severi and C. Villani, *Org. Biomol. Chem.*, 2016, **14**, 11137–11147, DOI: 10.1039/c6ob02145j.

(216) F. C. Tucci, T. Hu, M. F. Mesleh, A. Bokser, E. Allsopp, T. D. Gross, Z. Guo, Y.-F. Zhu, R. S. Struthers, N. Ling et al., *Chirality*, 2005, **17**, 559–564, DOI: 10.1002/chir.20198.

(217) L. D. Colebrook and H. G. Giles, *Can. J. Chem.*, 1975, **53**, 3431–3434, DOI: 10.1139/v75-491.

(218) A. Palani, S. Shapiro, J. W. Clader, W. J. Greenlee, D. Blythin, K. Cox, N. E. Wagner, J. Strizki, B. M. Baroudy and N. Dan, *Bioorg. Med. Chem. Lett.*, 2003, **13**, 705–708, DOI: 10.1016/s0960-894x(02)01062-4.

(219) Y. Shi, M.-H. Huang, J. E. Macor and D. E. Hughes, *J. Chromatogr. A*, 2005, **1078**, 67–73, DOI: 10.1016/j.chroma.2005.04.091.

(220) N. Mardirossian and M. Head-Gordon, *J. Chem. Phys.*, 2016, **144**, 214110, DOI: 10.1063/1.4952647.

(221) F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305, DOI: 10.1039/B508541A.

(222) S. Ehlert, M. Stahn, S. Spicher and S. Grimme, *J. Chem. Theory Comput.*, 2021, **17**, 4250–4261, DOI: 10.1021/acs.jctc.1c00471.

(223) S. Kulyk, S. M. De Paul, M. A. Marx, T. M. Peakman and C. R. Smith, *ACS Omega*, 2022, **7**, 32062–32067, DOI: 10.1021/acsomega.2c03316.

(224) Z. Wu, B. Ramsundar, E. N. Feinberg, J. Gomes, C. Geniesse, A. S. Pappu, K. Leswing and V. Pande, *Chem. Sci.*, 2018, **9**, 513–530, DOI: 10.1039/c7sc02664a.

(225) G. Subramanian, B. Ramsundar, V. Pande and R. A. Denny, *J. Chem. Inf. Model.*, 2016, **56**, 1936–1949, DOI: 10.1021/acs.jcim.6b00290.

(226) D. Butina, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 747–750, DOI: 10.1021/ci9803381.

(227) Q. Guo, S. Hernandez and P. Ballester, *ChemRxiv*, 2025, DOI: 10.26434/chemrxiv-2024-f1v2v-v2.

(228) F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and É. Duchesnay, *Scikit-learn: Machine Learning in Python*, 2018, DOI: 10.48550/arXiv.1201.0490.

(229) C. Zhao, R. Wu, S. Zhang and X. Hong, *J. Phys. Chem. A*, 2023, **127**, 6791–6803, DOI: 10.1021/acs.jpca.3c04215.

(230) P. Császár and P. Pulay, *J. Mol. Struct.*, 1984, **114**, 31–34, DOI: 10.1016/s0022-2860(84)87198-7.

(231) Ö. Farkas and H. B. Schlegel, *Phys. Chem. Chem. Phys.*, 2002, **4**, 11–15, DOI: 10.1039/b108658h.

(232) X. Li and M. J. Frisch, *J. Chem. Theory Comput.*, 2006, **2**, 835–839, DOI: 10.1021/ct050275a.

(233) A. Banerjee, N. Adams, J. Simons and R. Shepard, *J. Phys. Chem.*, 1985, **89**, 52–57, DOI: 10.1021/j100247a015.

(234) P. Pracht, S. Grimme, C. Bannwarth, F. Bohle, S. Ehlert, G. Feldmann, J. Gorges, M. Müller, T. Neudecker, C. Plett et al., *J. Chem. Phys.*, 2024, **160**, 114110, DOI: 10.1063/5.0197592.

(235) E. B. Wilson Jr, *J. Chem. Phys.*, 1941, **9**, 76–84, DOI: 10.1063/1.1750829.

(236) R. Lindh, A. Bernhardsson, G. Karlström and P.-Å. Malmqvist, *Chem. Phys. Lett.*, 1995, **241**, 423–428, DOI: 10.1016/0009-2614(95)00646-l.

(237) R. J. Gillespie, *Chem. Soc. Rev.*, 1992, **21**, 59–69, DOI: 10.1039/cs9922100059.

(238) T. A. Albright, *Acc. Chem. Res.*, 1982, **15**, 149–155, DOI: 10.1021/ar00077a004.

(239) D. Hernández-Valdés, F. Avignon, P. Müller, G. Meola, B. Probst, T. Fox, B. Spingler and R. Alberto, *Dalton Trans.*, 2020, **49**, 5250–5256, DOI: 10.1039/D0DT00731E.

(240) W. Kabsch, *Acta Crystallogr. Sect. A*, 1976, **32**, 922–923, DOI: 10.1107/s0567739476001873.

(241) E. R. Davidson, *J. Comput. Phys.*, 1975, **17**, 87–94, DOI: https://doi.org/10.1016/0021-9991(75)90065-0.

(242) Å. Björck, *Linear Algebra Appl.*, 1994, **197**, 297–316, DOI: 10.1016/0024-3795(94)90493-6.

(243) P. Pulay and G. Fogarasi, *J. Chem. Phys.*, 1992, **96**, 2856–2860, DOI: 10.1063/1.462844.

(244) J. Nocedal and S. J. Wright, in *Numerical Optimization*, See Section 3.1 for the backtracking line search algorithm and Armijo condition., Springer New York, New York, NY, 2006, p. 33, DOI: 10.1007/978-0-387-40065-5.

(245) M. O. Ishitsuka, T. Kusumi, H. Kakisawa, K. Kaya and M. M. Watanabe, *J. Am. Chem. Soc.*, 1990, **112**, 8180–8182, DOI: 10.1021/ja00178a060.

(246) J.-H. Chung and J.-D. Chai, *Sci. Rep.*, 2019, **9**, 2907, DOI: 10.1038/s41598-019-39524-4.

(247) J. Xu, Y. Li, Y. Xiang and X. Chen, *Nanoscale Res. Lett.*, 2013, **8**, 54, DOI: 10.1186/1556-276x-8-54.

(248) B. N. Parlett and I. S. Dhillon, *Linear Algebra Appl.*, 2000, **309**, 121–151, DOI: 10.1016/s0024-3795(99)00262-1.

(249) G. Millar, N. Weinberg and K. Mislow, *Mol. Phys.*, 2005, **103**, 2769–2772, DOI: 10.1080/00268970500217196.

(250) A. B. Buda and K. Mislow, *J. Am. Chem. Soc.*, 1992, **114**, 6006–6012, DOI: 10.1021/ja00041a016.

(251) A. B. Buda, T. A. der Heyde and K. Mislow, *Angew. Chem. Int. Ed.*, 1992, **31**, 989–1007, DOI: 10.1002/anie.199209891.

(252) A. Ferrarini and P. L. Nordio, *J. Chem. Soc., Perkin Trans. 2*, 1998, **2**, 455–460, DOI: 10.1039/a702619f.

(253) A. V. Luzanov and E. N. Babich, *J. Mol. Struct.: THEOCHEM*, 1995, **333**, 279–290, DOI: 10.1016/0166-1280(94)03944-g.

(254) M. Osipov, B. Pickup and D. Dunmur, *Mol. Phys.*, 1995, **84**, 1193–1206, DOI: 10.1080/00268979500100831.

(255) S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2013, **68**, 314–319, DOI: https://doi.org/10.1016/j.commatsci.2012.10.028.

(256) R. A. Sayle, *J. Comput. Aided Mol. Des.*, 2010, **24**, 485–496, DOI: 10.1007/s10822-010-9329-5.

(257) A. K. Colter and L. M. Clemens, *J. Phys. Chem.*, 1964, **68**, 651–654, DOI: 10.1021/j10078 5a503.

(258) G. Bringmann, S. Tasler, H. Endress, J. Kraus, K. Messer, M. Wohlfarth and W. Lobin, *J. Am. Chem. Soc.*, 2001, **123**, 2703–2711, DOI: 10.1021/ja003488c.

(259) M. Varela, C. Cabezas, J. C. Lopez and J. L. Alonso, *J. Phys. Chem. A*, 2013, **117**, 13275–13278, DOI: 10.1021/jp404581z.

(260) J. M. Goodman, I. Pletnev, P. Thiessen, E. Bolton and S. R. Heller, *J. Cheminformatics*, 2021, **13**, 40, DOI: 10.21203/rs.3.rs-378701/v1.

(261) A. Katbashev, M. Stahn, T. Rose, V. Alizadeh, M. Friede, C. Plett, P. Steinbach and S. Ehlert, *J. Phys. Chem. A*, 2025, **129**, 2667–2682, DOI: 10.1021/acs.jpca.4c08263.

(262) A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit and A. R. Leach, *Nuc. Acids Res.*, 2017, **45**, D945–D954, DOI: 10.1093/nar/gkw1074.

(263) A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Christensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schütt, M. Strange, K. S. Thygesen, T. Vegge, L. Vilhelmsen, M. Walter, Z. Zeng and K. W. Jacobsen, *J. Phys.: Condens. Matter*, 2017, **29**, 273002, DOI: 10.1088/1361-648x/aa680e.

(264) J. Nocedal and S. J. Wright, in *Numerical Optimization*, Springer New York, New York, NY, 2006, pp. 192–221, DOI: 10.1007/978-0-387-40065-5_6.

(265) V. Thönnißen, I. L. Atodiresei and F. W. Patureau, *Chem. Eur. J.*, 2023, **29**, e202300279, DOI: 10.1002/chem.202300780.

(266) J. Chandrasekhar, R. Dick, J. Van Veldhuizen, D. Koditek, E.-I. Lepist, M. E. McGrath, L. Patel, G. Phillips, K. Sedillo, J. R. Somoza et al., *J. Med. Chem.*, 2018, **61**, 6858–6868, DOI: 10.1021/acs.jmedchem.8b00797.

(267) R. J. Friary, M. Spangler, R. Osterman, L. Schulman and J. H. Schwerdt, *Chirality*, 1996, **8**, 364–371, DOI: 10.1002/(sici)1520-636x(1996)8:5<364::aid-chir2>3.0.co;2-e.

(268) S. Guile, J. Bantick, D. Cheshire, M. Cooper, A. Davis, D. Donald, R. Evans, C. Eyssade, D. Ferguson, S. Hill et al., *Bioorg. Med. Chem. Lett.*, 2006, **16**, 2260–2265, DOI: 10.1002/chin.200629210.

(269) Q.-S. Hu, D. Vitharana and L. Pu, *Tetrahedron Asymmetry*, 1995, **6**, 2123–2126, DOI: 10.1016/0957-4166(95)00280-3.

(270) M. S. Tichenor, J. J. Wiener, N. L. Rao, C. Pooley Deckhut, J. K. Barbay, K. D. Kreutter, G. M. Bacani, J. Wei, L. Chang, H. E. Murrey et al., *ACS Med. Chem. Lett.*, 2021, **12**, 782–790, DOI: 10.1021/acsmedchemlett.1c00044.

(271) A. L. Leivers, M. Tallant, J. B. Shotwell, S. Dickerson, M. R. Leivers, O. B. McDonald, J. Gobel, K. L. Creech, S. L. Strum, A. Mathis et al., *J. Med. Chem.*, 2014, **57**, 2091–2106, DOI: 10.1021/jm400781h.

(272) F. Satoh, S. Ito, H. Itoh, H. Rakugi, H. Shibata, A. Ichihara, M. Omura, K. Takahashi, Y. Okuda and S. Iijima, *Hypertens. Res.*, 2021, **44**, 464–472, DOI: 10.1038/s41440-020-00570-5.

(273) J. G. Kettle, S. K. Bagal, S. Bickerton, M. S. Bodnarchuk, J. Breed, R. J. Carbajo, D. J. Cassar, A. Chakraborty, S. Cosulich, I. Cumming et al., *J. Med. Chem.*, 2020, **63**, 4468–4483, DOI: 10.1021/acs.jmedchem.9b01720.

(274) Z. Szlávik, L. Ondi, M. Csékei, A. Paczal, Z. B. Szabó, G. Radics, J. Murray, J. Davidson, I. Chen, B. Davis et al., *J. Med. Chem.*, 2019, **62**, 6913–6924, DOI: 10.1021/acs.jmedchem.9b00134.

(275) J. M. Rohde, S. Karavadhi, R. Pragani, L. Liu, Y. Fang, W. Zhang, A. McIver, H. Zheng, Q. Liu, M. I. Davis et al., *J. Med. Chem.*, 2021, **64**, 4913–4946, DOI: 10.1021/acs.jmedchem.1c00019.

(276) J. Chandrasekhar, R. Dick, J. Van Veldhuizen, D. Koditek, E.-I. Lepist, M. E. McGrath, L. Patel, G. Phillips, K. Sedillo, J. R. Somoza et al., *J. Med. Chem.*, 2018, **61**, 6858–6868, DOI: 10.1021/acs.jmedchem.8b00797.

(277) K. Hoegenauer, J. Kallen, E. Jimenez-Nunez, R. Strang, P. Ertl, N. G. Cooke, S. Hintermann, M. Voegtle, C. Betschart, D. J. McKay et al., *J. Med. Chem.*, 2019, **62**, 10816–10832, DOI: 10.1021/acs.jmedchem.9b01291.

(278) P. P. Patwardhan, K. S. Ivy, E. Musi, E. de Stanchina and G. K. Schwartz, *Oncotarget*, 2015, **7**, 4093, DOI: 10.18632/oncotarget.6547.

(279) Y. Sun, J. A. Alberta, C. Pilarz, D. Calligaris, E. J. Chadwick, S. H. Ramkissoon, L. A. Ramkissoon, V. M. Garcia, E. Mazzola, L. Goumnerova et al., *Neuro Oncol.*, 2017, **19**, 774–785, DOI: 10.1093/neuonc/now261.

(280) G. Bringmann and D. Menche, *Angew. Chem. Int. Ed.*, 2001, **40**, 1687–1690, DOI: 10.1002/1521-3773(20010504)40:9<1687::aid-anie16870>3.3.co;2-y.

(281) P. Cheng, D. L. Clive, S. Fernandopulle and Z. Chen, *Chem. Commun.*, 2013, **49**, 558–560, DOI: 10.1039/c2cc37110c.

(282) S. Cen, N. Huang, D. Lian, A. Shen, M.-X. Zhao and Z. Zhang, *Nat. Commun.*, 2022, **13**, 4735, DOI: 10.1038/s41467-022-32432-8.

(283) G. Yang, S. Sun, Z. Li, Y. Liu and J. Wang, *Commun. Chem.*, 2021, **4**, 144, DOI: 10.1038/s42004-021-00580-5.

(284) L. Xing, B. Devadas, R. V. Devraj, S. R. Selness, H. Shieh, J. K. Walker, M. Mao, D. Messing, B. Samas, J. Z. Yang et al., *ChemMedChem*, 2012, **7**, 273–280, DOI: 10.1002/cmdc.201100439.

(285) J. Wang, W. Zeng, S. Li, L. Shen, Z. Gu, Y. Zhang, J. Li, S. Chen and X. Jia, *ACS Med. Chem. Lett.*, 2017, **8**, 299–303, DOI: 10.1021/acsmedchemlett.6b00465.

(286) M. D. Shelley, L. Hartley, R. G. Fish, P. Groundwater, J. J. Morgan, D. Mort, M. Mason and A. Evans, *Cancer Lett.*, 1999, **135**, 171–180, DOI: 10.1016/S0304-3835(98)00302-4.

(287) N. A. Stamos, B. Ries, R. Schneider, P. Tzvetkova, F. Montel, C. Jandl and U. Werthmann, *Mol. Pharmaceutics*, 2025, DOI: 10.26434/chemrxiv-2025-rz4q6.

(288) A. Zask, J. Murphy and G. A. Ellestad, *J. Med. Chem.*, 2016, **59**, 5770–5782, DOI: 10.1021/acs.jmedchem.6b01088.

(289) S. M. Hoy, *Drugs*, 2022, **82**, 1671–1679, DOI: 10.1007/s40265-022-01796-y.

(290) R. L. Ferris, K. Harrington, J. D. Schoenfeld, M. Tahara, C. Esdar, S. Salmio, A. Schroeder and J. Bourhis, *Cancer Treat. Rev.*, 2023, **113**, 102492, DOI: 10.1016/j.ctrv.2022.102492.

(291) A. W. Hird and et al., *J. Med. Chem.*, 2020, **63**, 13562–13579, DOI: 10.1021/acs.jmedchem.0c01234.

(292) J. E. Bradner, A. Monkhouse and et al., *ACS Med. Chem. Lett.*, 2023, **14**, 1181–1188, DOI: 10.1021/acsmedchemlett.3c00123.

(293) J. Bergman, A. Parsons and et al., *Mol. Cancer Ther.*, 2024, **23**, 1535–1546, DOI: 10.1158/1535-7163.MCT-22-0241.

(294) C. Deur, A. K. Agrawal, H. Baum, J. Booth, S. Bove, J. Brieland, A. Bunker, C. Connolly, J. Cornicelli, J. Dumin et al., *Bioorg. Med. Chem. Lett.*, 2007, **17**, 4599–4603, DOI: 10.1016/j.bmcl.2007.05.083.

(295) J. Porter, A. Payne, I. Whitcombe, B. de Candole, D. Ford, R. Garlish, A. Hold, B. Hutchinson, G. Trevitt, J. Turner et al., *Bioorg. Med. Chem. Lett.*, 2009, **19**, 1767–1772, DOI: 10.1016/j.bmcl.2009.01.071.

(296) J. S. Albert, D. Aharony, D. Andisik, H. Barthlow, P. R. Bernstein, R. A. Bialecki, R. Dedinas, B. T. Dembofsky, D. Hill, K. Kirkland et al., *J. Med. Chem.*, 2002, **45**, 3972–3983, DOI: 10.1021/jm020094i.

(297) A. E. Tron, M. A. Belmonte, A. Adam, B. M. Aquila, L. H. Boise, E. Chiarparin, J. Cidado, K. J. Embrey, E. Gangl, F. D. Gibbons et al., *Nat. Commun.*, 2018, **9**, 5341, DOI: 10.1038/s41467-018-07551-w.

(298) J. J. Crawford, A. R. Johnson, D. L. Misner, L. D. Belmont, G. Castanedo, R. Choy, M. Coraggio, L. Dong, C. Eigenbrot, R. Erickson, N. Ghilardi, J. Hau, A. Katewa, P. B. Kohli, W. Lee, J. W. Lubach, B. S. McKenzie, D. F. Ortwine, L. Schutt, S. Tay, B. Wei, K. Reif, L. Liu, H. Wong and W. B. Young, *J. Med. Chem.*, 2018, **61**, PMID: 29457982, 2227–2245, DOI: 10.1021/acs.jmedchem.7b01712.

(299) K. Nicolaou, N. L. Simmons, J. S. Chen, N. M. Haste and V. Nizet, *Tetrahedron Lett.*, 2011, **52**, 2041–2043, DOI: 10.1016/j.tetlet.2010.09.059.

(300) A. P. Degnan and A. Meyers, *J. Am. Chem. Soc.*, 1999, **121**, 2762–2769, DOI: 10.1021/ja984182x.

(301) a. Miyashita, A. Yasuda, H. Takaya, K. Toriumi, T. Ito, T. Souchi and R. Noyori, *J. Am. Chem. Soc.*, 1980, **102**, 7932–7934, DOI: 10.1021/ja00547a020.

(302) J. Clayden, S. P. Fletcher, J. J. McDouall and S. J. Rowbottom, *J. Am. Chem. Soc.*, 2009, **131**, 5331–5343, DOI: 10.1021/ja900722q.

(303) P. Eveleigh, E. C. Hulme, C. Schudt and N. J. M. Birdsall, *Mol. Pharmacol.*, 1989, **35**, 477–483, DOI: 10.1016/s0026-895x(25)11262-5.

(304) J. R. Healy, S. A. Smith and et al., *J. Med. Chem.*, 2019, **62**, 3428–3440, DOI: 10.1021/acs.jmedchem.8b01831.

(305) F. A. L. M. Eskens, M. J. A. de Jonge, P. Bhargava, T. Isoe, M. M. Cotreau, B. Esteves, K. Hayashi, H. Burger, M. Thomeer, L. van Doorn and J. Verweij, *Clin. Cancer Res.*, 2011, **17**, 7156–7163, DOI: 10.1158/1078-0432.CCR-11-0411.

(306) A. R. Todd, *Q. Rev., Chem. Soc.*, 1949, **3**, 230–252, DOI: 10.1039/QR9490300230.

List of Mathematical Symbols

Symbol	Description
$\mathbf{A}_{\text{adj}}$	Adjacency matrix
a_{ij}	Matrix element of adjacency matrix
a_1, a_2	Empirical parameters for steric repulsion energy
b	Bias term
b_{ij}^{Coulomb}	Coulomb matrix elements
$b_{ij}^{\text{topography}}$	Topography matrix elements
$\mathbf{B}$	Wilson B-matrix
$\mathbf{B}^{\text{topography}}$	Topography matrix
BO_{ij}	Bond order matrix
$c_{\mu k}$	Coefficient of basis function μ in orbital k
$\mathbf{C}$	Matrix of eigenvectors (MO coefficients)
C^s	Set of chemical rings
CN_i	Coordination number of atom i
$\tilde{\text{CN}}_i$	D3 coordination number
$\mathbf{D}$	Graph distance matrix
d_{ij}	Number of edges in shortest path between i and j
$\deg(i)$	Degree of vertex i
e_{ij}	Edge feature between nodes i and j
E	Total energy of the molecule
E_{elec}	Electronic energy
$E_{\text{rep}}^{(m)}$	Steric repulsion energy of conformation m
$E(G)$	Set of edges in molecular graph G
$\mathbf{E}$	Diagonal matrix of eigenvalues
F_{edge}	Edge feature dimension
F_{node}	Node feature dimension
$g_{ab}^{\text{chirality}}$	Chirality matrix elements
$\mathbf{G}^{\text{chirality}}$	Chirality matrix
$\boldsymbol{\mathcal{G}}$	Gyration tensor
G	Molecular graph
G_0	Osipov-Pickup-Dunmur chirality measure
G_a	Chirality index for fragment a
$\mathbf{g}_0$	Gradient vector at reference geometry

h_{ij}^{topology}	Topology matrix elements
$h_i^{(l)}$	Feature vector of node i at layer l
$\hat{H}$	Hamiltonian operator
$\hat{H}_{\text{elec}}$	Electronic Hamiltonian operator
$\mathbf{H}^{\text{topology}}$	Topology matrix
$\mathbf{H}$	Hessian matrix in Cartesian coordinates
$\mathbf{H}_Q$	Hessian matrix in internal coordinates
$\mathcal{H}$	Hidden dimension
i, j	Atom/Electron indices
(i, j)	Bond between atoms i and j
$\boldsymbol{\mathcal{I}}$	Rigidity matrix
k	Orbital index
$k_1^{\text{D3}}, k_2^{\text{D3}}$	D3 fit parameter
l	Layer index in the neural network
$\mathcal{L}$	Loss function
m_i	Mass of particle i
M^l	Message function at layer l
$M_{\text{electronic}}$	Set of electronically hindered bonds
n_1, n_2	Integer exponents in chirality index
n_{atoms}	Number of atoms in the molecule
n_{basis}	Number of basis functions
n_{bonds}	Number of bonds in the molecule
n_{conf}	Number of conformations sampled
n_{elec}	Number of electrons
n_{modes}	Number of vibrational modes after excluding soft modes
n_{orb}	Number of orbitals
n_{data}	Number of training samples
n_k	Occupation number of orbital k
$N(i)$	Set of neighboring atoms of atom i
$N_{v,i}$	Valence of atom i
$\mathbf{N}_{\text{occ}}$	Occupation matrix
p_a	Priority of fragment a
p_i	Priority of atom i
p_{ij}^{steric}	Probability that bond (i, j) is sterically hindered
p_{ij}^{WBO}	Probability that bond (i, j) is electronically hindered
$\mathbf{P}$	Projection matrix for rigid-body modes
$\mathbf{P}^{\text{dens}}$	Density matrix
$\mathbf{P}^{\text{MolBar}}$	MolBar density matrix
q_i	Partial charge of atom i
q_{mol}	Total charge of the molecule

Q	Internal coordinates of the molecule
$\mathbf{r}$	Electronic/particle coordinates
r_{ij}	Euclidean distance between atoms i and j
r_i^{cov}	Covalent radius of atom i
r_{ij}^{vdW}	Sum of van der Waals radii of atoms i and j
$\mathbf{r}_{ij}$	Distance vector between atoms i and j
$\mathbf{R}$	Cartesian coordinates of the molecule
$\mathbf{R}_0$	Reference Cartesian coordinates
$\Delta\mathbf{R}$	Displacement from reference geometry in Cartesian coordinates
$\mathbf{R}_a^{\text{uni}}$	Unified/idealized Cartesian coordinates of fragment a
$\mathbf{S}$	Overlap matrix
t	Time
$t_{1/2}$	Half-life time
$\hat{T}$	Kinetic energy operator
$\hat{T}_{\text{elec}}$	Electronic kinetic energy operator
$\hat{T}_{\text{nuc}}$	Nuclear kinetic energy operator
u_i	Unsaturation degree of atom i
U^l	Update function at layer l
$\mathbf{U}$	Eigenvectors of the Hessian matrix
v	Vibrational quantum number
$\hat{V}$	Potential energy operator
$\hat{V}_{ij}$	Potential energy operator between particles i and j
$\hat{V}_{\text{elec}-\text{elec}}$	Electron-electron potential energy operator
$\hat{V}_{\text{elec}-\text{nuc}}$	Electron-nucleus potential energy operator
$\hat{V}_{\text{nuc}-\text{nuc}}$	Nucleus-nucleus potential energy operator
$V(G)$	Set of nodes in molecular graph G
W^*	Optimal weight parameters
x	Feature vector of the molecule
y	Predicted target property
$\hat{y}$	Ground truth target property
Z_i	Atomic number of atom i
α_{ijk}	Angle formed by atoms i, j, and k
α_{ijk}^0	Reference angle formed by atoms i, j, and k
ϵ_0	Vacuum permittivity
θ_{ijkl}	Dihedral angle formed by atoms i, j, k, and l
θ_{ijkl}^0	Reference dihedral angle formed by atoms i, j, k, and l
κ	Force constant for harmonic potentials
λ	Parameter controlling step size in RFO method
μ, ν	Basis function indices
∇^2	Laplace operator

ξ	Step size parameter
$\rho(\mathbf{r}_i)$	Electron density at position $\mathbf{r}_i$
ρ_i	Artificial electron density at atom i
σ	Activation function
τ	Trust radius for step size control
$\phi(t)$	Temporal wave function component
ϕ_k	Molecular orbital k
ϕ_μ, ϕ_ν	Basis functions μ and ν
$\chi_{\text{nuc}}(\mathbf{r}_{\text{nuc}})$	Nuclear wave function
$\psi(\mathbf{r})$	Spatial wave function
$\psi_{\text{elec}}(\mathbf{r}; \mathbf{r}_{\text{nuc}})$	Electronic wave function
$\Psi(\mathbf{r}, t)$	Time-dependent wave function
ω_{steric}	Steric threshold for classification
ω_{WBO}	WBO threshold for classification
h	Planck constant
$\hbar$	Reduced Planck constant

Acronyms

2D	two-dimensional
3D	three-dimensional
ALPB	analytical linearized Poisson-Boltzmann
ANC	approximate normal coordinate
ANCOpt	approximate normal coordinate optimizer
aRST	automated reaction search tool
BACE	beta-secretase 1
BFGS	Broyden-Fletcher-Goldfarb-Shanno
BFS	breadth-first search
CC	coupled cluster
CI	configuration interaction
CIF	crystallographic information file
CML	chemical markup language
CN	coordination number
CNDO	complete neglect of differential overlap
CREST	conformer-rotamer-ensemble sampling tool
CT	chemical table
Ctab	connectivity table
CTD	connect-the-dots
CV	cross-validation
DFT	density functional theory
DFTB	density functional tight binding
DMPNN	directed message passing neural network
DNMR	dynamic nuclear magnetic resonance
EEQ	electronegativity equilibration
EGNN	equivariant graph neural network
EHT	extended Hückel theory
ELF	electron localization function
GDIIS	geometry direct inversion in the iterative subspace
GNN	graph neural network

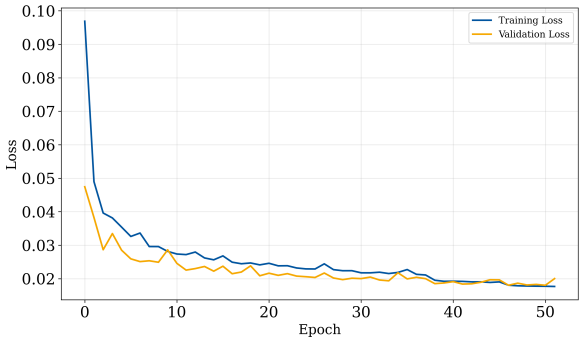
GPR	Gaussian process regression
g-xTB	general-purpose extended tight-binding
GFN2-xTB	geometries, frequencies, and noncovalent interactions - extended tight-binding
GOAT	global optimization algorithm
HIP-NN	hierarchically interacting particle neural network
HPLC	high-performance liquid chromatography
InChI	international chemical identifier
INDO	intermediate neglect of differential overlap
IRC	intrinsic reaction coordinate
IUPAC	International Union of Pure and Applied Chemistry
LCAO	linear combination of atomic orbitals
MD	molecular dynamics
MDL	Molecular Design Limited
MERP	minimum energy reaction pathway
MINDO	modified intermediate neglect of differential overlap
ML	machine learning
MLIP	machine-learning interatomic potential
MO	molecular orbital
MolBar	Molecular Barcode
MolBarOpt	MolBar optimizer
MoLE	mixture of linear experts
Molfile	molecular file
MPNN	message passing neural network
MSE	mean squared error
NBO	natural bond orbitals
NDDO	neglect of diatomic differential overlap
NIST	National Institute of Standards and Technology
NMR	nuclear magnetic resonance
PDA	probability density analysis
PDB	Protein Data Bank
PES	potential energy surface
PhysNet	physical neural network
QMugs	quantum-mechanical properties of drug-like molecules

QSAR	quantitative structure-activity relationship
QTAIM	quantum theory of atoms in molecules
ReLU	rectified linear unit
RFO	rational function optimization
RMSD	root-mean-square deviation
ROSDAL	representation of structure diagram arranged linearly
RRHO	rigid-rotor harmonic-oscillator
SCF	self-consistent field
SDfile	structure-data file
SELFIES	self-referencing embedded strings
SLN	SYBYL line notation
SMILES	simplified molecular-input line-entry system
SQM	semiempirical quantum mechanics
tmQM	transition metal quantum mechanics
TTConf	tensor train conformer search
UMA	universal models for atoms
UMAP	uniform manifold approximation and projection
VB	valence bond
vdW	van der Waals
VSEPR	valence shell electron pair repulsion
WBO	Wiberg bond order
WLN	Wiswesser line-formula notation
YAeHMOP	yet another extended Hückel molecular orbital package
ZPE	zero-point energy

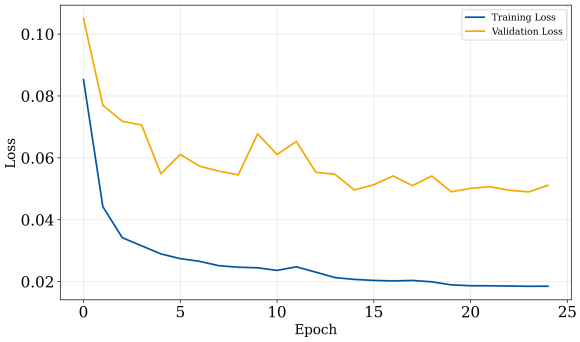
A Figures

A.1 Electronic Model: Cross-Validation Folds

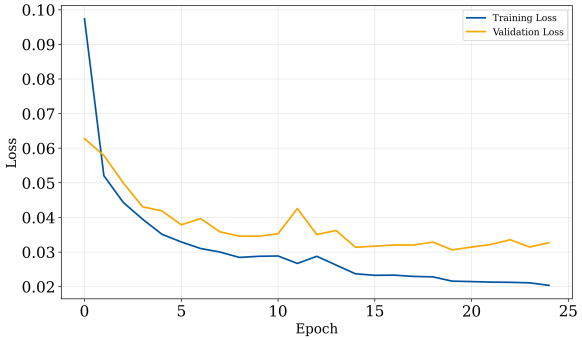
A stratified 5-fold cross-validation was conducted to evaluate the performance and generalizability of the electronic model. The learning curves for each cross-validation fold are illustrated in Figure A.1a to A.1f. Fold 4 shows a significant spike in validation loss at epoch 4, which immediately returns to the previous level in the subsequent epoch, suggesting a transient overfitting event or an outlier in the training data for that fold. However, the model quickly recovers and continues to converge, ultimately achieving strong performance on the validation set. The other folds exhibit smooth convergence without significant fluctuations, indicating stable training across different data partitions.



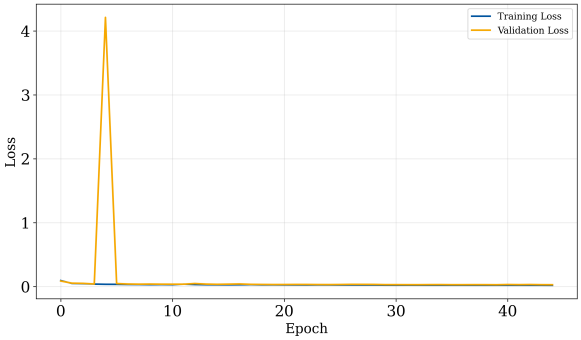
(a) Cross-validation fold 1



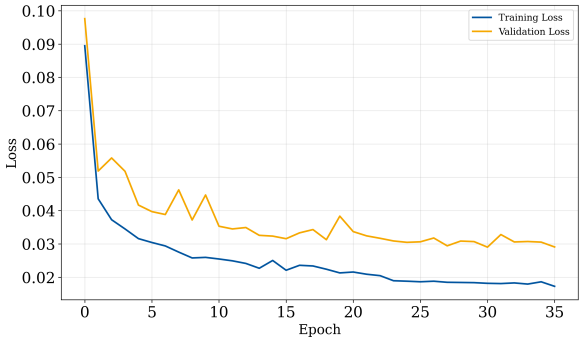
(b) Cross-validation fold 2



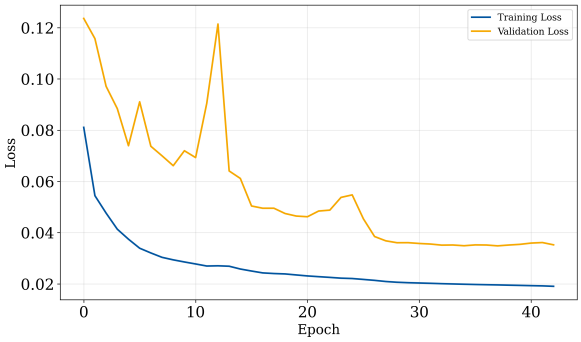
(c) Cross-validation fold 3



(d) Cross-validation fold 4



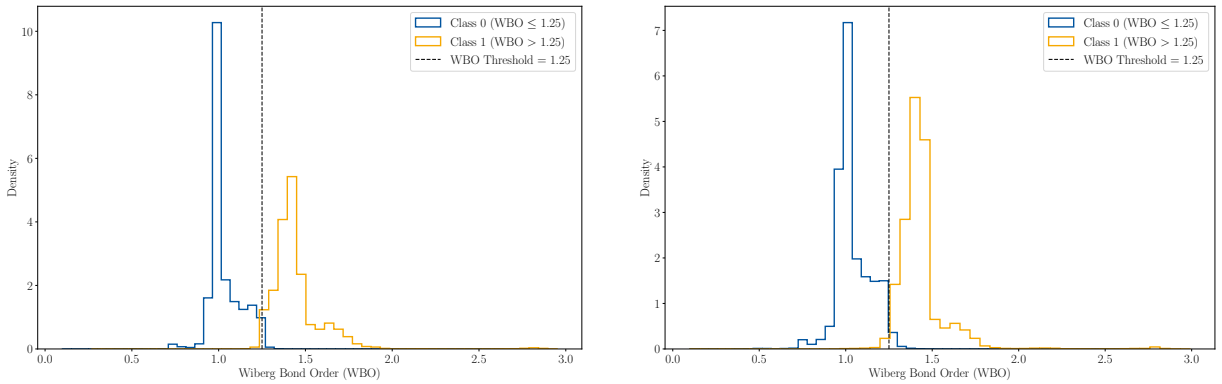
(e) Cross-validation fold 5



(f) Final model

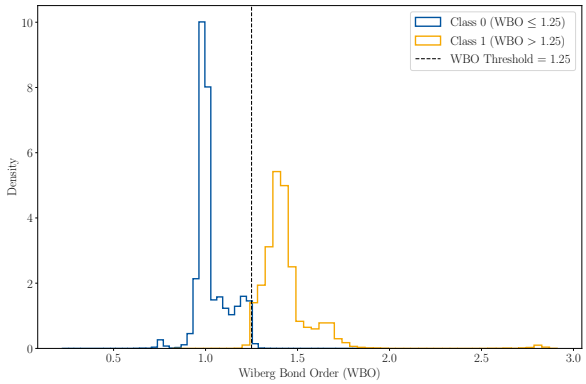
Figure A.1: Learning curves for the electronic model for each cross-validation fold and the final model, showing training and validation loss over epochs. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

Figures A.2a to A.2e present the distribution of predicted WBOs for the test sets in each cross-validation fold. The classifier effectively distinguishes between electronically hindered bonds (orange) and flexible bonds (blue). Most hindered bonds show a reference WBO above 1.25, while non-hindered bonds are predominantly below this threshold. A minor overlap around the 1.25 threshold exists for each fold, which remains negligible relative to the overall classification performance.

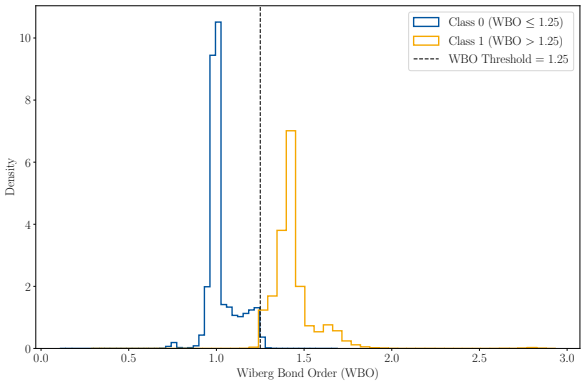


(a) Cross-validation fold 1

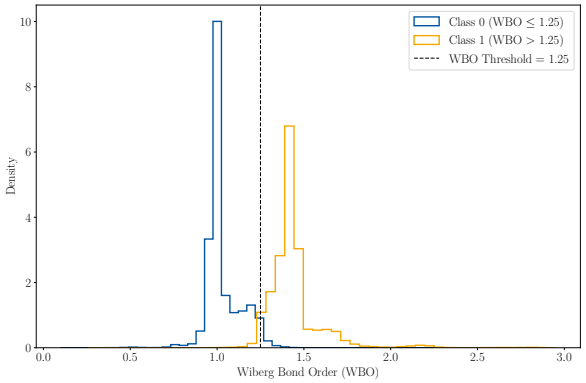
(b) Cross-validation fold 2



(c) Cross-validation fold 3



(d) Cross-validation fold 4



(e) Cross-validation fold 5

Figure A.2: Distribution of WBO values predicted by the electronic hindrance classifier for the test set in each cross-validation fold. The x-axis represents the WBO, and the y-axis indicates the number of bonds. The classifier demonstrates a clear separation between bonds predicted as electronically hindered (orange) and non-hindered (blue). Most electronically hindered bonds exhibit WBO values exceeding 1.25, whereas non-hindered bonds typically show values below this threshold. The slight overlap near 1.25 reflects both the model's predictive uncertainty and the conformational dependence of the WBO in the reference data. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

A.2 Steric Model: Dataset

Table B.1: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

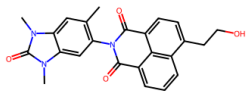
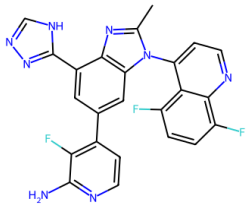
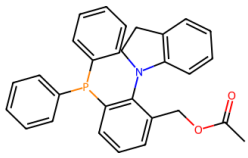
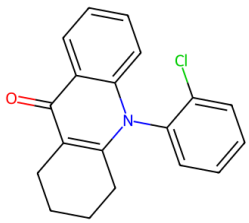
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	12	32.2	[265]
	13	34.5	[266]
	14	32.1	[265]
	19	36.4	[267]

Table B.2: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

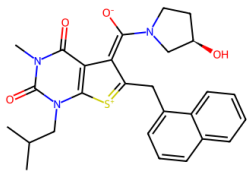
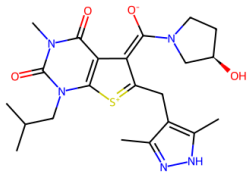
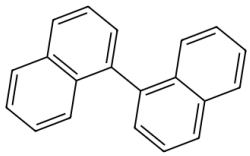
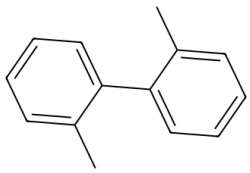
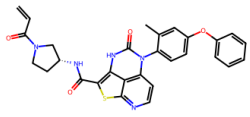
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	22	20.5	[268]
	24	22.3	[268]
	27	35.5	[269] (derivative of 28)
	29	22.2	[119]
	30	35.5	[270]

Table B.3: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

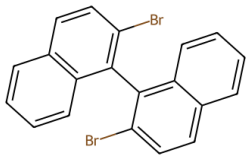
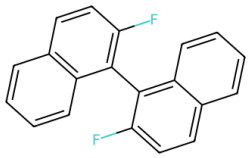
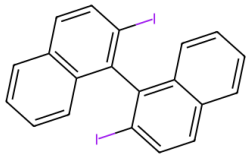
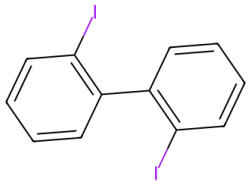
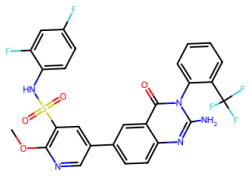
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	31	57.8	[269] (derivative of 28)
	35	40.4	[269] (derivative of 28)
	38	61.3	[269] (derivative of 28)
	39	26.9	[119]
	40	38.3	[271]

Table B.4: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

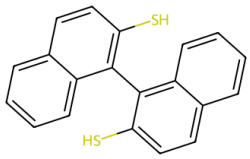
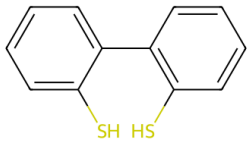
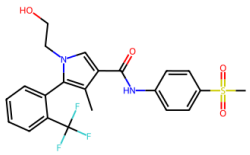
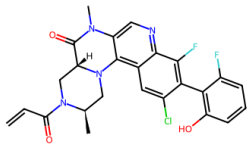
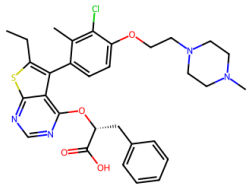
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	41	48.7	[269] (derivative of 28)
	42	21.3	[119]
	43	28.6	[272]
	45	36.8	[273]
	48	37.4	[274]

Table B.5: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

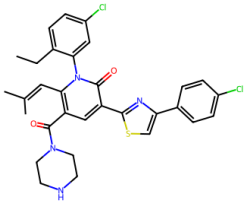
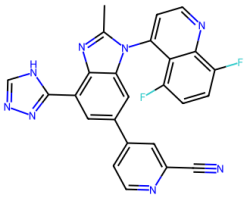
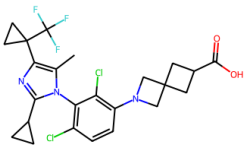
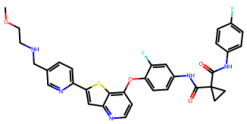
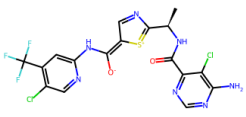
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	49	40.8	[275]
	51	34.5	[276]
	52	55.7	[277]
	53	22.7	[278]
	54	22.7	[279]

Table B.6: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

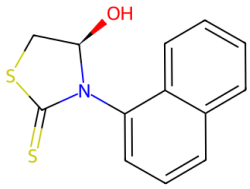
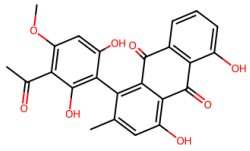
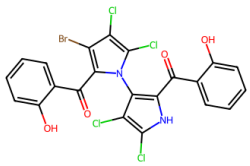
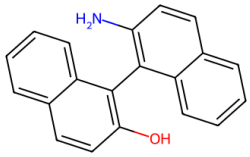
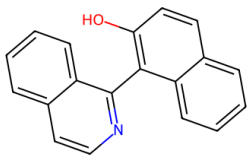
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	55	25.2	[215]
	57	39.0	[280]
	59	42.2	[281]
	62	47.4	[282]
	64	32.4	[283]

Table B.7: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

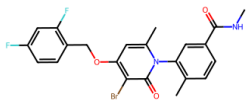
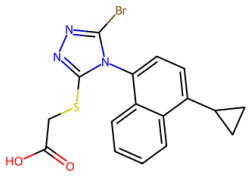
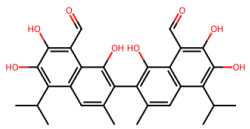
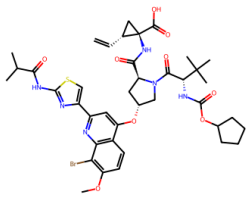
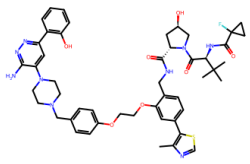
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	65	37.6	[284]
	66	36.7	[285]
	68	45.9	[286]
	69	22.7	[287]
	70	22.7	[287]

Table B.8: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

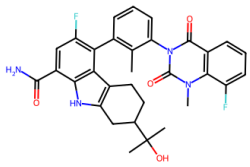
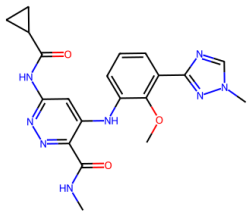
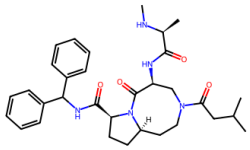
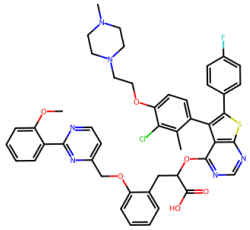
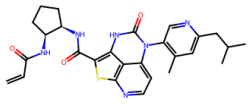
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	72	34.9	[288]
	75	22.7	[289]
	76	20.5	[290]
	77	37.4	[291]
	78	32.6	[292]

Table B.9: Added atropisomeric compounds known from literature with their computed rotational barrier for the training and validation dataset. Only the largest rotational barrier is shown for each compound.

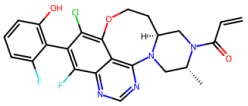
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	79	38.1	[293]

Table B.10: Atropisomeric compounds that did not converge during TS optimization involving the atropisomeric axis known from the literature were added to the training and validation datasets, along with their computed rotational barriers.

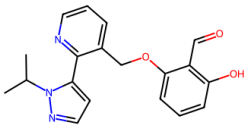
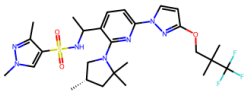
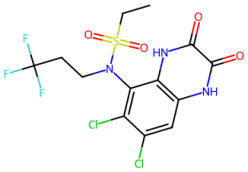
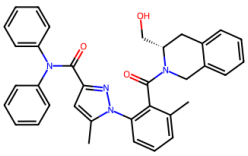
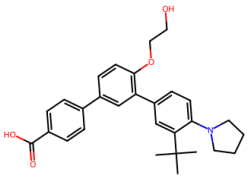
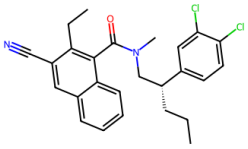
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	15	N/A	[198]
	16	N/A	[198]
	17	N/A	[294]
	18	N/A	[295]
	20	N/A	[198]
	21	N/A	[296]

Table B.11: Atropisomeric compounds that did not converge during TS optimization involving the atropisomeric axis known from the literature were added to the training and validation datasets, along with their computed rotational barriers.

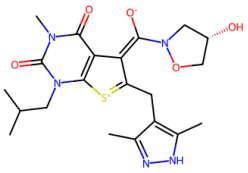
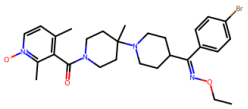
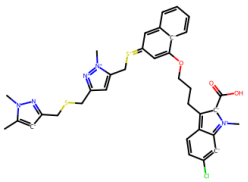
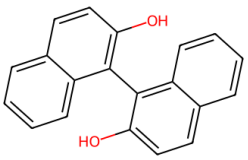
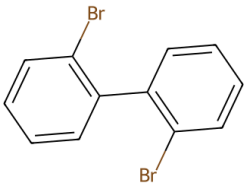
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	23	N/A	[268]
	25	N/A	[218]
	26	N/A	[297]
	28	N/A	[269]
	32	N/A	[119]

Table B.12: Atropisomeric compounds that did not converge during TS optimization involving the atropisomeric axis known from the literature were added to the training and validation datasets, along with their computed rotational barriers.

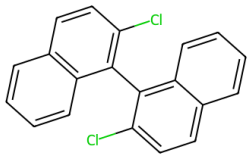
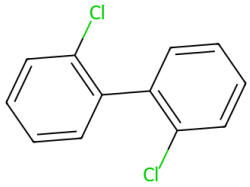
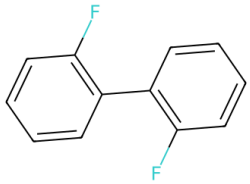
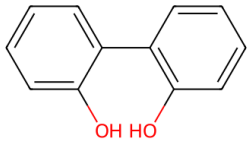
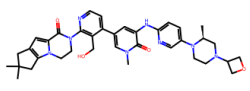
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	33	N/A	[269] (derivative of 28)
	34	N/A	[119]
	36	N/A	[119]
	37	N/A	[119]
	44	N/A	[298]

Table B.13: Atropisomeric compounds that did not converge during TS optimization involving the atropisomeric axis known from the literature were added to the training and validation datasets, along with their computed rotational barriers.

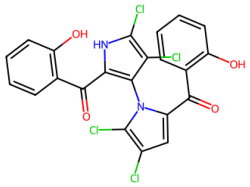
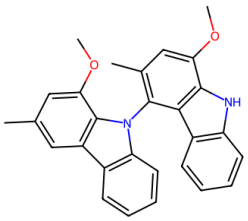
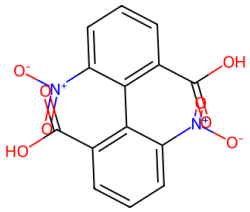
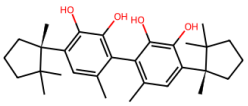
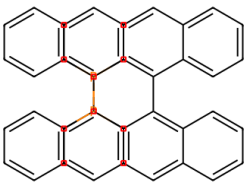
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	47	N/A	[299]
	50	N/A	[258]
	56	N/A	[199]
	58	N/A	[300]
	60	N/A	[301]

Table B.14: Atropisomeric compounds that did not converge during TS optimization involving the atropisomeric axis known from the literature were added to the training and validation datasets, along with their computed rotational barriers.

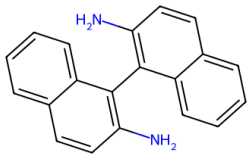
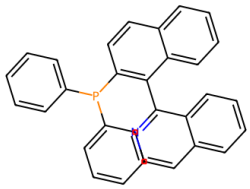
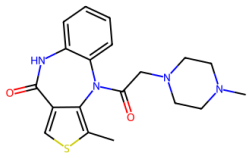
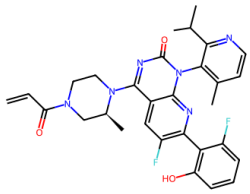
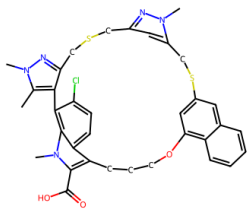
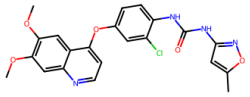
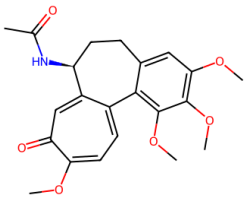
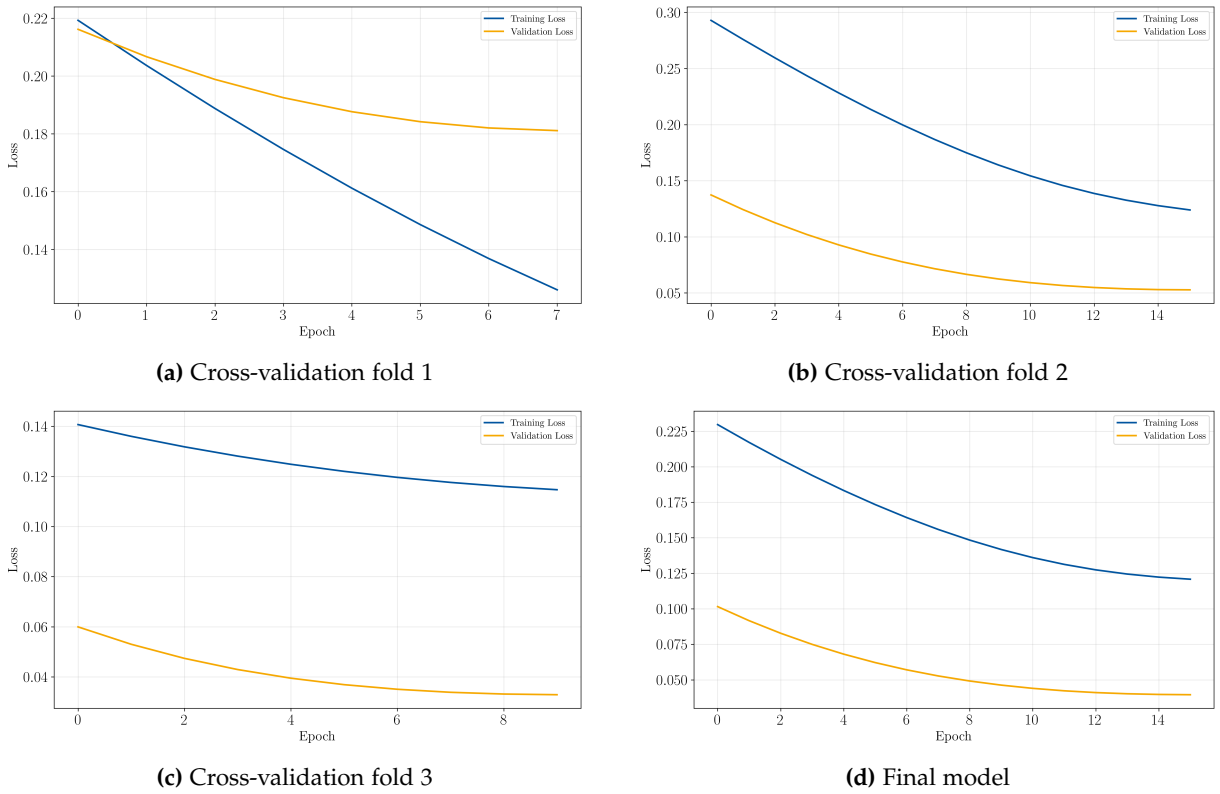
Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	61	N/A	[282]
	63	N/A	[302]
	67	N/A	[303]
	71	N/A	[203]
	73	N/A	[304]

Table B.15: Atropisomeric compounds that did not converge during TS optimization involving the atropisomeric axis known from the literature were added to the training and validation datasets, along with their computed rotational barriers.

Structure	Molecule (Name)	$\Delta G^{\ddagger}_{\text{calc.}}$	Source
	74	N/A	[305]
	80	N/A	[306]

A.3 Steric Model: Cross-Validation Folds

The learning curves for each cross-validation fold are shown in Figure A.3a to A.3d. Training was terminated early in all folds once the validation loss stabilized, signaling convergence.



(a) Cross-validation fold 1

(b) Cross-validation fold 2

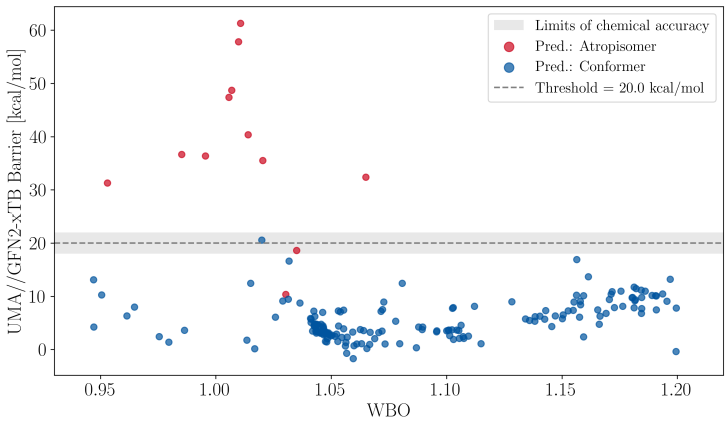
(c) Cross-validation fold 3

(d) Final model

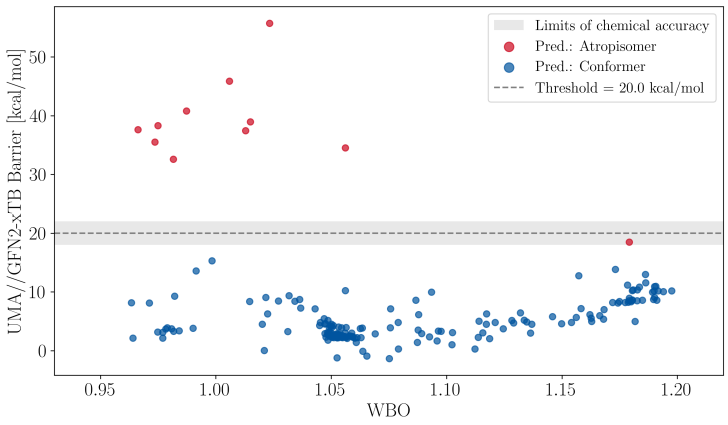
Figure A.3: Learning curves for training steric model for each cross-validation and final model showing training and validation loss over epochs. Validation loss is lower than training loss due to the use of weighted binary cross-entropy loss to address class imbalance, which assigns higher weights to the minority class in the training set, leading to a different effective loss landscape compared to the validation set where weights are not applied. Early stopping was applied to prevent overfitting, with training terminating once the validation loss plateaued for five consecutive epochs. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

To assess model performance across the stratified cross-validation, Figure A.4a to A.4c present the predicted classification results for each of the three validation folds individually. For each fold, the computed rotational barrier is plotted against the Wiberg bond order (WBO) of the central bond. Points are color-coded by predicted class, with atropisomeric predictions shown in red and conformer predictions in blue. A horizontal dashed line at 20kcal/mol represents the conventional threshold for atropisomeric stability. A grey shaded region (18–22kcal/mol) indicates the range of chemical accuracy, where DFT-level barrier predictions are particularly uncertain. In all folds, the classifier reliably identifies atropisomeric fragments, with false predictions occurring primarily within the grey uncertainty region. These ambiguous fragments often have barriers near the 20 kcal/mol threshold, making their classification challenging even

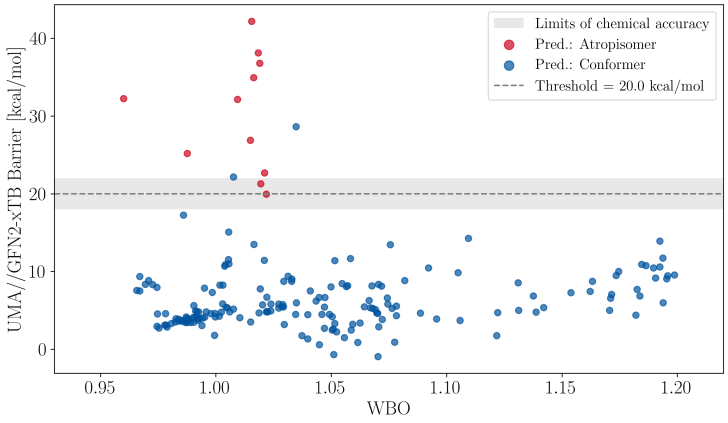
with high-level quantum chemical methods.



(a) Validation fold 1



(b) Validation fold 2



(c) Validation fold 3

Figure A.4: Model predictions for the test set in each cross-validation fold. The computed rotational barriers are plotted against the Wiberg bond order (WBO) of the central bond. Red points denote fragments predicted as atropisomeric, while blue points indicate non-atropisomeric predictions. The dashed line at 20kcal/mol signifies the classification threshold, and the grey shaded area (18–22kcal/mol) highlights the region of DFT uncertainty or chemical ambiguity. Most misclassifications fall within this range, underscoring the model's sensitivity to borderline cases. Reproduced with permission from: Nils van Staalduinen and Christoph Bannwarth, *J. Chem. Inf. Model.* **2026**, 66, 10935–10955. DOI: `10.1021/acs.jcim.6c00688`. Copyright 2026 American Chemical Society.

B First and Second Order Derivatives of the Structure Unification Force Field

The structure unification workflow uses a force field optimization to determine the ideal geometry of a fragment. For this, a simple harmonic force field is defined to constrain bond lengths, angles, and dihedrals to reference values extracted from the input structures. The energy function is given by

$$\begin{aligned} E_{\text{FF}} =& E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{Coulomb}} \\ =& \sum E_{\text{bond},ij} + \sum E_{\text{angle},ijk} + \sum E_{\text{dihedral},ijkl} + \sum E_{\text{Coulomb},ij} \end{aligned} \tag{.1}$$

Bond Potential

One single bond harmonic potential between two atoms i and j is given by

$$E_{\text{bond},ij} = k_{\text{bond}} \left(r_{ij}^0 - r_{ij} \right)^2 \ , \tag{.2}$$

where distance in Cartesian space between two atoms i and j reads

$$r_{ij} = \sqrt{x_{ij}^2 + y_{ij}^2 + z_{ij}^2} \tag{.3}$$

with

$$x_{ij} = x_i - x_j \tag{.4}$$

and x_i is the x-coordinate of atom i.

Derivatives of $E_{\text{bond},ij}$

The first and second order derivatives of $E_{\text{bond},ij}$ with respect to the coordinates x, y and z of the participating atoms i and j are:

$$\frac{d}{dx_i} E_{\text{bond},ij} = 2k_{\text{bond}} \left(r_{ij} - r_{ij}^0 \right) \frac{d}{dx_i} r_{ij} \tag{.5}$$

$$\frac{d}{dy_i} E_{\text{bond},ij} = 2k_{\text{bond}} \left(r_{ij} - r_{ij}^0 \right) \frac{d}{dy_i} r_{ij} \tag{.6}$$

$$\frac{d}{dz_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d}{dz_i}r_{ij} \tag{.7}$$

$$\frac{d}{dx_j}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d}{dx_j}r_{ij} \tag{.8}$$

$$\frac{d}{dy_j}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d}{dy_j}r_{ij} \tag{.9}$$

$$\frac{d}{dz_j}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d}{dz_j}r_{ij} \tag{.10}$$

$$\frac{d^2}{dx_i^2}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dx_i^2}r_{ij} + 2k_{\text{bond}}\left(\frac{d}{dx_i}r_{ij}\right)^2 \tag{.11}$$

$$\frac{d^2}{dy_i dx_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dy_i dx_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_i}r_{ij}\frac{d}{dy_i}r_{ij} \tag{.12}$$

$$\frac{d^2}{dz_i dx_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_i dx_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_i}r_{ij}\frac{d}{dz_i}r_{ij} \tag{.13}$$

$$\frac{d^2}{dx_j dx_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dx_j dx_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_i}r_{ij}\frac{d}{dx_j}r_{ij} \tag{.14}$$

$$\frac{d^2}{dy_j dx_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dy_j dx_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_i}r_{ij}\frac{d}{dy_j}r_{ij} \tag{.15}$$

$$\frac{d^2}{dz_j dx_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_j dx_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_i}r_{ij}\frac{d}{dz_j}r_{ij} \tag{.16}$$

$$\frac{d^2}{dy_i^2}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dy_i^2}r_{ij} + 2k_{\text{bond}}\left(\frac{d}{dy_i}r_{ij}\right)^2 \tag{.17}$$

$$\frac{d^2}{dz_i dy_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_i dy_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dy_i}r_{ij}\frac{d}{dz_i}r_{ij} \tag{.18}$$

$$\frac{d^2}{dx_j dy_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dy_i dx_j}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_j}r_{ij}\frac{d}{dy_i}r_{ij} \tag{.19}$$

$$\frac{d^2}{dy_j dy_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dy_j dy_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dy_i}r_{ij}\frac{d}{dy_j}r_{ij} \tag{.20}$$

$$\frac{d^2}{dz_j dy_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_j dy_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dy_i}r_{ij}\frac{d}{dz_j}r_{ij} \tag{.21}$$

$$\frac{d^2}{dz_i^2}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_i^2}r_{ij} + 2k_{\text{bond}}\left(\frac{d}{dz_i}r_{ij}\right)^2 \tag{.22}$$

$$\frac{d^2}{dx_j dz_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_i dx_j}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_j}r_{ij}\frac{d}{dz_i}r_{ij} \tag{.23}$$

$$\frac{d^2}{dy_j dz_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_i dy_j}r_{ij} + 2k_{\text{bond}}\frac{d}{dy_j}r_{ij}\frac{d}{dz_i}r_{ij} \tag{.24}$$

$$\frac{d^2}{dz_j dz_i}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_j dz_i}r_{ij} + 2k_{\text{bond}}\frac{d}{dz_i}r_{ij}\frac{d}{dz_j}r_{ij} \tag{.25}$$

$$\frac{d^2}{dx_j^2}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dx_j^2}r_{ij} + 2k_{\text{bond}}\left(\frac{d}{dx_j}r_{ij}\right)^2 \tag{.26}$$

$$\frac{d^2}{dy_j dx_j}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dy_j dx_j}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_j}r_{ij}\frac{d}{dy_j}r_{ij} \tag{.27}$$

$$\frac{d^2}{dz_j dx_j}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_j dx_j}r_{ij} + 2k_{\text{bond}}\frac{d}{dx_j}r_{ij}\frac{d}{dz_j}r_{ij} \tag{.28}$$

$$\frac{d^2}{dy_j^2}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dy_j^2}r_{ij} + 2k_{\text{bond}}\left(\frac{d}{dy_j}r_{ij}\right)^2 \tag{.29}$$

$$\frac{d^2}{dz_j dy_j}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_j dy_j}r_{ij} + 2k_{\text{bond}}\frac{d}{dy_j}r_{ij}\frac{d}{dz_j}r_{ij} \tag{.30}$$

$$\frac{d^2}{dz_j^2}E_{\text{bond},ij} = 2k_{\text{bond}}\left(r_{ij} - r_{ij}^0\right)\frac{d^2}{dz_j^2}r_{ij} + 2k_{\text{bond}}\left(\frac{d}{dz_j}r_{ij}\right)^2 \tag{.31}$$

The derivatives of r_{ij} with respect to the coordinates of the participating atoms i and j are given in Section B.

Derivatives of r_{ij}

The first and second order derivatives of r_{ij} with respect to the coordinates x, y and z of the participating atoms i and j are:

$$\frac{d}{dx_i}r_{ij} = -\frac{x_{ji}}{r_{ij}} \tag{.32}$$

$$\frac{d}{dy_i}r_{ij} = -\frac{y_{ji}}{r_{ij}} \tag{.33}$$

$$\frac{d}{dz_i}r_{ij} = -\frac{z_{ji}}{r_{ij}} \tag{.34}$$

$$\frac{d}{dx_j}r_{ij} = \frac{x_{ji}}{r_{ij}} \tag{.35}$$

$$\frac{d}{dy_j}r_{ij} = \frac{y_{ji}}{r_{ij}} \tag{.36}$$

$$\frac{d}{dz_j}r_{ij} = \frac{z_{ji}}{r_{ij}} \tag{.37}$$

$$\frac{d^2}{dx_i^2}r_{ij} = \frac{1}{r_{ij}} - \frac{x_{ji}^2}{r_{ij}^3} \tag{.38}$$

$$\frac{d^2}{dy_i dx_i}r_{ij} = -\frac{x_{ji}y_{ji}}{r_{ij}^3} \tag{.39}$$

$$\frac{d^2}{dz_i dx_i}r_{ij} = -\frac{x_{ji}z_{ji}}{r_{ij}^3} \tag{.40}$$

$$\frac{d^2}{dx_j dx_i} r_{ij} = -\frac{1}{r_{ij}} + \frac{x_{ji}^2}{r_{ij}^3} \tag{.41}$$

$$\frac{d^2}{dy_j dx_i} r_{ij} = \frac{x_{ji} y_{ji}}{r_{ij}^3} \tag{.42}$$

$$\frac{d^2}{dz_j dx_i} r_{ij} = \frac{x_{ji} z_{ji}}{r_{ij}^3} \tag{.43}$$

$$\frac{d^2}{dy_i^2} r_{ij} = \frac{1}{r_{ij}} - \frac{y_{ji}^2}{r_{ij}^3} \tag{.44}$$

$$\frac{d^2}{dz_i dy_i} r_{ij} = -\frac{y_{ji} z_{ji}}{r_{ij}^3} \tag{.45}$$

$$\frac{d^2}{dx_j dy_i} r_{ij} = \frac{x_{ji} y_{ji}}{r_{ij}^3} \tag{.46}$$

$$\frac{d^2}{dy_j dy_i} r_{ij} = -\frac{1}{r_{ij}} + \frac{y_{ji}^2}{r_{ij}^3} \tag{.47}$$

$$\frac{d^2}{dz_j dy_i} r_{ij} = \frac{y_{ji} z_{ji}}{r_{ij}^3} \tag{.48}$$

$$\frac{d^2}{dz_i^2} r_{ij} = \frac{1}{r_{ij}} - \frac{z_{ji}^2}{r_{ij}^3} \tag{.49}$$

$$\frac{d^2}{dx_j dz_i} r_{ij} = \frac{x_{ji} z_{ji}}{r_{ij}^3} \tag{.50}$$

$$\frac{d^2}{dy_j dz_i} r_{ij} = \frac{y_{ji} z_{ji}}{r_{ij}^3} \tag{.51}$$

$$\frac{d^2}{dz_j dz_i} r_{ij} = -\frac{1}{r_{ij}} + \frac{z_{ji}^2}{r_{ij}^3} \tag{.52}$$

$$\frac{d^2}{dx_j^2} r_{ij} = \frac{1}{r_{ij}} - \frac{x_{ji}^2}{r_{ij}^3} \tag{.53}$$

$$\frac{d^2}{dy_j dx_j} r_{ij} = -\frac{x_{ji} y_{ji}}{r_{ij}^3} \tag{.54}$$

$$\frac{d^2}{dz_j dx_j} r_{ij} = -\frac{x_{ji} z_{ji}}{r_{ij}^3} \tag{.55}$$

$$\frac{d^2}{dy_j^2} r_{ij} = \frac{1}{r_{ij}} - \frac{y_{ji}^2}{r_{ij}^3} \tag{.56}$$

$$\frac{d^2}{dz_j dy_j} r_{ij} = -\frac{y_{ji} z_{ji}}{r_{ij}^3} \tag{.57}$$

$$\frac{d^2}{dz_j^2} r_{ij} = \frac{1}{r_{ij}} - \frac{z_{ji}^2}{r_{ij}^3} \tag{.58}$$

The first and second order derivatives of r_{ij} are equivalent to the derivatives of e.g. r_{ji} by interchanging atomic indices.

Angle Potential

The angle harmonic potential between three atoms i, j and k is given by

$$E_{\text{angle},ijk} = k_{\text{angle}} \left(\alpha_{ijk}^0 - \alpha_{ijk}\right)^2 \tag{.59}$$

The angle between three atoms i, j and k is given by

$$\alpha_{ijk} = \arccos\Big(\frac{x_{ij}x_{kj} + y_{ij}y_{kj} + z_{ij}z_{kj}}{r_{ij}r_{kj}}\Big) \tag{.60}$$

$$= \arccos\Big(\frac{a_{ijk}}{r_{ij}r_{kj}}\Big)\,, \tag{.61}$$

where a_{ijk} is the dot product of the vectors $\mathbf{r}_{ij}$ and $\mathbf{r}_{kj}$.

Derivatives of $E_{\text{angle},ijk}$

The first and second order derivatives of $E_{\text{angle},ijk}$ with respect to the coordinates x, y and z of the participating atoms i, j and k are:

$$\frac{d}{dx_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha_{ijk}^0\right) \frac{d}{dx_i} \alpha_{ijk} \tag{.62}$$

$$\frac{d}{dy_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha_{ijk}^0\right) \frac{d}{dy_i} \alpha_{ijk} \tag{.63}$$

$$\frac{d}{dz_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha_{ijk}^0\right) \frac{d}{dz_i} \alpha_{ijk} \tag{.64}$$

$$\frac{d}{dx_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha_{ijk}^0\right) \frac{d}{dx_j} \alpha_{ijk} \tag{.65}$$

$$\frac{d}{dy_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha_{ijk}^0\right) \frac{d}{dy_j} \alpha_{ijk} \tag{.66}$$

$$\frac{d^2}{dx_i^2}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dx_i^2}\alpha_{ijk} + 2k_{\text{angle}}\left(\frac{d}{dx_i}\alpha_{ijk}\right)^2 \tag{.71}$$

$$\frac{d^2}{dy_i dx_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dy_i dx_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_i}\alpha_{ijk}\frac{d}{dy_i}\alpha_{ijk} \tag{.72}$$

$$\frac{d^2}{dz_i dx_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dz_i dx_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_i}\alpha_{ijk}\frac{d}{dz_i}\alpha_{ijk} \tag{.73}$$

$$\frac{d^2}{dx_j dx_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dx_j dx_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_i}\alpha_{ijk}\frac{d}{dx_j}\alpha_{ijk} \tag{.74}$$

$$\frac{d^2}{dy_j dx_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dy_j dx_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_i}\alpha_{ijk}\frac{d}{dy_j}\alpha_{ijk} \tag{.75}$$

$$\frac{d^2}{dz_j dx_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dz_j dx_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_i}\alpha_{ijk}\frac{d}{dz_j}\alpha_{ijk} \tag{.76}$$

$$\frac{d^2}{dx_k dx_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dx_k dx_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_i}\alpha_{ijk}\frac{d}{dx_k}\alpha_{ijk} \tag{.77}$$

$$\frac{d^2}{dy_k dx_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dy_k dx_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_i}\alpha_{ijk}\frac{d}{dy_k}\alpha_{ijk} \tag{.78}$$

$$\frac{d^2}{dz_k dx_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dz_k dx_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_i}\alpha_{ijk}\frac{d}{dz_k}\alpha_{ijk} \tag{.79}$$

$$\frac{d^2}{dy_i^2}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dy_i^2}\alpha_{ijk} + 2k_{\text{angle}}\left(\frac{d}{dy_i}\alpha_{ijk}\right)^2 \tag{.80}$$

$$\frac{d^2}{dz_i dy_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dz_i dy_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dy_i}\alpha_{ijk}\frac{d}{dz_i}\alpha_{ijk} \tag{.81}$$

$$\frac{d^2}{dx_j dy_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dy_i dx_j}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dx_j}\alpha_{ijk}\frac{d}{dy_i}\alpha_{ijk} \tag{.82}$$

$$\frac{d^2}{dy_j dy_i}E_{\text{angle},ijk} = 2k_{\text{angle}}\left(\alpha_{ijk} - \alpha_{ijk}^0\right)\frac{d^2}{dy_j dy_i}\alpha_{ijk} + 2k_{\text{angle}}\frac{d}{dy_i}\alpha_{ijk}\frac{d}{dy_j}\alpha_{ijk} \tag{.83}$$

$$\frac{d^2}{dz_j dy_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_j dy_i}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_i}\alpha_{ijk} \frac{d}{dz_j}\alpha_{ijk} \tag{.84}$$

$$\frac{d^2}{dx_k dy_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_i dx_k}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_k}\alpha_{ijk} \frac{d}{dy_i}\alpha_{ijk} \tag{.85}$$

$$\frac{d^2}{dy_k dy_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_k dy_i}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_i}\alpha_{ijk} \frac{d}{dy_k}\alpha_{ijk} \tag{.86}$$

$$\frac{d^2}{dz_k dy_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_k dy_i}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_i}\alpha_{ijk} \frac{d}{dz_k}\alpha_{ijk} \tag{.87}$$

$$\frac{d^2}{dz_i^2} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_i^2}\alpha_{ijk} + 2k_{\text{angle}} \left(\frac{d}{dz_i}\alpha_{ijk}\right)^2 \tag{.88}$$

$$\frac{d^2}{dx_j dz_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_i dx_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_j}\alpha_{ijk} \frac{d}{dz_i}\alpha_{ijk} \tag{.89}$$

$$\frac{d^2}{dy_j dz_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_i dy_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_j}\alpha_{ijk} \frac{d}{dz_i}\alpha_{ijk} \tag{.90}$$

$$\frac{d^2}{dz_j dz_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_j dz_i}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dz_i}\alpha_{ijk} \frac{d}{dz_j}\alpha_{ijk} \tag{.91}$$

$$\frac{d^2}{dx_k dz_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_i dx_k}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_k}\alpha_{ijk} \frac{d}{dz_i}\alpha_{ijk} \tag{.92}$$

$$\frac{d^2}{dy_k dz_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_i dy_k}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_k}\alpha_{ijk} \frac{d}{dz_i}\alpha_{ijk} \tag{.93}$$

$$\frac{d^2}{dz_k dz_i} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_k dz_i}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dz_i}\alpha_{ijk} \frac{d}{dz_k}\alpha_{ijk} \tag{.94}$$

$$\frac{d^2}{dx_j^2} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dx_j^2}\alpha_{ijk} + 2k_{\text{angle}} \left(\frac{d}{dx_j}\alpha_{ijk}\right)^2 \tag{.95}$$

$$\frac{d^2}{dy_j dx_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_j dx_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_j}\alpha_{ijk} \frac{d}{dy_j}\alpha_{ijk} \tag{.96}$$

$$\frac{d^2}{dz_j dx_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_j dx_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_j}\alpha_{ijk} \frac{d}{dz_j}\alpha_{ijk} \tag{.97}$$

$$\frac{d^2}{dx_k dx_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dx_k dx_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_j}\alpha_{ijk} \frac{d}{dx_k}\alpha_{ijk} \tag{.98}$$

$$\frac{d^2}{dy_k dx_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_k dx_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_j}\alpha_{ijk} \frac{d}{dy_k}\alpha_{ijk} \tag{.99}$$

$$\frac{d^2}{dz_k dx_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_k dx_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_j}\alpha_{ijk} \frac{d}{dz_k}\alpha_{ijk} \tag{.100}$$

$$\frac{d^2}{dy_j^2} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_j^2}\alpha_{ijk} + 2k_{\text{angle}} \left(\frac{d}{dy_j}\alpha_{ijk}\right)^2 \tag{.101}$$

$$\frac{d^2}{dz_j dy_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_j dy_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_j}\alpha_{ijk} \frac{d}{dz_j}\alpha_{ijk} \tag{.102}$$

$$\frac{d^2}{dx_k dy_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_j dx_k}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_k}\alpha_{ijk} \frac{d}{dy_j}\alpha_{ijk} \tag{.103}$$

$$\frac{d^2}{dy_k dy_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_k dy_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_j}\alpha_{ijk} \frac{d}{dy_k}\alpha_{ijk} \tag{.104}$$

$$\frac{d^2}{dz_k dy_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_k dy_j}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_j}\alpha_{ijk} \frac{d}{dz_k}\alpha_{ijk} \tag{.105}$$

$$\frac{d^2}{dz_j^2} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_j^2}\alpha_{ijk} + 2k_{\text{angle}} \left(\frac{d}{dz_j}\alpha_{ijk}\right)^2 \tag{.106}$$

$$\frac{d^2}{dx_k dz_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_j dx_k}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_k}\alpha_{ijk} \frac{d}{dz_j}\alpha_{ijk} \tag{.107}$$

$$\frac{d^2}{dy_k dz_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_j dy_k}\alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_k}\alpha_{ijk} \frac{d}{dz_j}\alpha_{ijk} \tag{.108}$$

$$\frac{d^2}{dz_k dz_j} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_k dz_j} \alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dz_j} \alpha_{ijk} \frac{d}{dz_k} \alpha_{ijk} \tag{.109}$$

$$\frac{d^2}{dx_k^2} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dx_k^2} \alpha_{ijk} + 2k_{\text{angle}} \left(\frac{d}{dx_k} \alpha_{ijk}\right)^2 \tag{.110}$$

$$\frac{d^2}{dy_k dx_k} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_k dx_k} \alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_k} \alpha_{ijk} \frac{d}{dy_k} \alpha_{ijk} \tag{.111}$$

$$\frac{d^2}{dz_k dx_k} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_k dx_k} \alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dx_k} \alpha_{ijk} \frac{d}{dz_k} \alpha_{ijk} \tag{.112}$$

$$\frac{d^2}{dy_k^2} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dy_k^2} \alpha_{ijk} + 2k_{\text{angle}} \left(\frac{d}{dy_k} \alpha_{ijk}\right)^2 \tag{.113}$$

$$\frac{d^2}{dz_k dy_k} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_k dy_k} \alpha_{ijk} + 2k_{\text{angle}} \frac{d}{dy_k} \alpha_{ijk} \frac{d}{dz_k} \alpha_{ijk} \tag{.114}$$

$$\frac{d^2}{dz_k^2} E_{\text{angle},ijk} = 2k_{\text{angle}} \left(\alpha_{ijk} - \alpha^0_{ijk}\right) \frac{d^2}{dz_k^2} \alpha_{ijk} + 2k_{\text{angle}} \left(\frac{d}{dz_k} \alpha_{ijk}\right)^2 \tag{.115}$$

The derivatives of α_{ijk} with respect to the coordinates x, y and z of the participating atoms i, j and k are defined in Section B.

Derivatives of α_{ijk}

The first and second order derivatives of α_{ijk} with respect to the coordinates x, y and z of the participating atoms i, j and k are:

$$\frac{d}{dx_i} \alpha_{ijk} = \frac{a_{ijk} \frac{d}{dx_i} r_{ij} - r_{ij} \frac{d}{dx_i} a_{ijk}}{r_{ij}^2 r_{kj} \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} \tag{.116}$$

$$\frac{d}{dy_i}\alpha_{ijk} = \frac{a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.117}$$

$$\frac{d}{dz_i}\alpha_{ijk} = \frac{a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.118}$$

$$\frac{d}{dx_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d}{dx_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dx_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dx_j}a_{ijk}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.119}$$

$$\frac{d}{dy_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.120}$$

$$\frac{d}{dz_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dz_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dz_j}a_{ijk}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.121}$$

$$\frac{d}{dx_k}\alpha_{ijk} = \frac{a_{ijk}\frac{d}{dx_k}r_{kj} - r_{kj}\frac{d}{dx_k}a_{ijk}}{r_{ij} r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.122}$$

$$\frac{d}{dy_k}\alpha_{ijk} = \frac{a_{ijk}\frac{d}{dy_k}r_{kj} - r_{kj}\frac{d}{dy_k}a_{ijk}}{r_{ij}r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2r_{kj}^2}+1}} \tag{.123}$$

$$\frac{d}{dz_k}\alpha_{ijk} = \frac{a_{ijk}\frac{d}{dz_k}r_{kj} - r_{kj}\frac{d}{dz_k}a_{ijk}}{r_{ij}r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2r_{kj}^2}+1}} \tag{.124}$$

$$\frac{d^2}{dx_i^2}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dx_i}r_{ij}}{r_{ij}^3r_{kj}^2} + \frac{a_{ijk}\frac{d}{dx_i}a_{ijk}}{r_{ij}^2r_{kj}^2}\right)}{r_{ij}^2r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dx_i^2}r_{ij} - r_{ij}\frac{d^2}{dx_i^2}a_{ijk}}{r_{ij}^2r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dx_i}r_{ij}}{r_{ij}^3r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2r_{kj}^2}+1}} \tag{.125}$$

$$\begin{aligned}\frac{d^2}{dy_idx_i}\alpha_{ijk} =& \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_i}r_{ij}}{r_{ij}^3r_{kj}^2} + \frac{a_{ijk}\frac{d}{dy_i}a_{ijk}}{r_{ij}^2r_{kj}^2}\right)}{r_{ij}^2r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dy_idx_i}r_{ij} - r_{ij}\frac{d^2}{dy_idx_i}a_{ijk} - \frac{d}{dx_i}a_{ijk}\frac{d}{dy_i}r_{ij} + \frac{d}{dy_i}a_{ijk}\frac{d}{dx_i}r_{ij}}{r_{ij}^2r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2r_{kj}^2}+1}} \\ &- \frac{2\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dy_i}r_{ij}}{r_{ij}^3r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2r_{kj}^2}+1}}\end{aligned} \tag{.126}$$

$$\frac{d^2}{dz_i dx_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_i}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dz_i}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_i dx_i}r_{ij} - r_{ij}\frac{d^2}{dz_i dx_i}a_{ijk} - \frac{d}{dx_i}a_{ijk}\frac{d}{dz_i}r_{ij} + \frac{d}{dz_i}a_{ijk}\frac{d}{dx_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.127}$$

$$- \frac{2\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dz_i}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dx_j dx_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dx_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dx_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dx_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dx_j dx_i}r_{ij} - r_{ij}\frac{d^2}{dx_j dx_i}a_{ijk} - \frac{d}{dx_i}a_{ijk}\frac{d}{dx_j}r_{ij} + \frac{d}{dx_j}a_{ijk}\frac{d}{dx_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.128}$$

$$- \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dx_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dx_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dy_j dx_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dy_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dy_j dx_i}r_{ij} - r_{ij}\frac{d^2}{dy_j dx_i}a_{ijk} - \frac{d}{dx_i}a_{ijk}\frac{d}{dy_j}r_{ij} + \frac{d}{dy_j}a_{ijk}\frac{d}{dx_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.129}$$

$$- \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dz_j dx_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dz_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_j dx_i}r_{ij} - r_{ij}\frac{d^2}{dz_j dx_i}a_{ijk} - \frac{d}{dx_i}a_{ijk}\frac{d}{dz_j}r_{ij} + \frac{d}{dz_j}a_{ijk}\frac{d}{dx_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.130}$$

$$- \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dx_k dx_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dx_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dx_k dx_i}a_{ijk} + \frac{d}{dx_k}a_{ijk}\frac{d}{dx_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.131}$$

$$\frac{d^2}{dy_k dx_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dy_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dy_k dx_i}a_{ijk} + \frac{d}{dy_k}a_{ijk}\frac{d}{dx_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.132}$$

$$\frac{d^2}{dz_k dx_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dz_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dz_k dx_i}a_{ijk} + \frac{d}{dz_k}a_{ijk}\frac{d}{dx_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dx_i}r_{ij} - r_{ij}\frac{d}{dx_i}a_{ijk}\right)\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.133}$$

$$\frac{d^2}{dy_i^2}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_i}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dy_i}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dy_i^2}r_{ij} - r_{ij}\frac{d^2}{dy_i^2}a_{ijk}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dy_i}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.134}$$

$$\frac{d^2}{dz_i dy_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_i}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dz_i}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_i dy_i}r_{ij} - r_{ij}\frac{d^2}{dz_i dy_i}a_{ijk} - \frac{d}{dy_i}a_{ijk}\frac{d}{dz_i}r_{ij} + \frac{d}{dz_i}a_{ijk}\frac{d}{dy_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dz_i}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.135}$$

$$\frac{d^2}{dx_j dy_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dx_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dx_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dx_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dy_i dx_j}r_{ij} - r_{ij}\frac{d^2}{dy_i dx_j}a_{ijk} + \frac{d}{dx_j}a_{ijk}\frac{d}{dy_i}r_{ij} - \frac{d}{dy_i}a_{ijk}\frac{d}{dx_j}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.136}$$

$$- \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dx_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dx_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dy_j dy_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dy_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dy_j dy_i}r_{ij} - r_{ij}\frac{d^2}{dy_j dy_i}a_{ijk} - \frac{d}{dy_i}a_{ijk}\frac{d}{dy_j}r_{ij} + \frac{d}{dy_j}a_{ijk}\frac{d}{dy_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.137}$$

$$- \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dz_j dy_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dz_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_j dy_i}r_{ij} - r_{ij}\frac{d^2}{dz_j dy_i}a_{ijk} - \frac{d}{dy_i}a_{ijk}\frac{d}{dz_j}r_{ij} + \frac{d}{dz_j}a_{ijk}\frac{d}{dy_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.138}$$

$$- \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dx_k dy_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dx_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dy_i dx_k}a_{ijk} + \frac{d}{dx_k}a_{ijk}\frac{d}{dy_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.139}$$

$$\frac{d^2}{dy_k dy_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dy_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dy_k dy_i}a_{ijk} + \frac{d}{dy_k}a_{ijk}\frac{d}{dy_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.140}$$

$$\frac{d^2}{dz_k dy_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dz_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dz_k dy_i}a_{ijk} + \frac{d}{dz_k}a_{ijk}\frac{d}{dy_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dy_i}r_{ij} - r_{ij}\frac{d}{dy_i}a_{ijk}\right)\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.141}$$

$$\frac{d^2}{dz_i^2}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_i}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dz_i}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_i^2}r_{ij} - r_{ij}\frac{d^2}{dz_i^2}a_{ijk}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dz_i}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.142}$$

$$\frac{d^2}{dx_j dz_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dx_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dx_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dx_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_i dx_j}r_{ij} - r_{ij}\frac{d^2}{dz_i dx_j}a_{ijk} + \frac{d}{dx_j}a_{ijk}\frac{d}{dz_i}r_{ij} - \frac{d}{dz_i}a_{ijk}\frac{d}{dx_j}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.143}$$

$$- \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dx_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dx_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dy_j dz_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dy_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_i dy_j}r_{ij} - r_{ij}\frac{d^2}{dz_i dy_j}a_{ijk} + \frac{d}{dy_j}a_{ijk}\frac{d}{dz_i}r_{ij} - \frac{d}{dz_i}a_{ijk}\frac{d}{dy_j}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.144}$$

$$- \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dz_j dz_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dz_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_j dz_i}r_{ij} - r_{ij}\frac{d^2}{dz_j dz_i}a_{ijk} - \frac{d}{dz_i}a_{ijk}\frac{d}{dz_j}r_{ij} + \frac{d}{dz_j}a_{ijk}\frac{d}{dz_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.145}$$

$$- \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dx_k dz_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dx_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dz_i dx_k}a_{ijk} + \frac{d}{dx_k}a_{ijk}\frac{d}{dz_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.146}$$

$$\frac{d^2}{dy_k dz_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dy_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dz_i dy_k}a_{ijk} + \frac{d}{dy_k}a_{ijk}\frac{d}{dz_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.147}$$

$$\frac{d^2}{dz_k dz_i}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dz_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{-r_{ij}\frac{d^2}{dz_k dz_i}a_{ijk} + \frac{d}{dz_k}a_{ijk}\frac{d}{dz_i}r_{ij}}{r_{ij}^2 r_{kj}\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{\left(a_{ijk}\frac{d}{dz_i}r_{ij} - r_{ij}\frac{d}{dz_i}a_{ijk}\right)\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.148}$$

$$\frac{d^2}{dx_j^2}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dx_j^2}r_{kj} + a_{ijk}r_{kj}\frac{d^2}{dx_j^2}r_{ij} + 2a_{ijk}\frac{d}{dx_j}r_{ij}\frac{d}{dx_j}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dx_j^2}a_{ijk}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.149}$$
$$+ \frac{\left(a_{ijk}r_{ij}\frac{d}{dx_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dx_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dx_j}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dx_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dx_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dx_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}}$$
$$- \frac{2\left(a_{ijk}r_{ij}\frac{d}{dx_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dx_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dx_j}a_{ijk}\right)\frac{d}{dx_j}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}r_{ij}\frac{d}{dx_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dx_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dx_j}a_{ijk}\right)\frac{d}{dx_j}r_{ij}}{r_{ij}^3 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dy_j dx_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dy_j dx_j}r_{kj} + a_{ijk}r_{kj}\frac{d^2}{dy_j dx_j}r_{ij} + a_{ijk}\frac{d}{dx_j}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}\frac{d}{dy_j}r_{ij}\frac{d}{dx_j}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dy_j dx_j}a_{ijk}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}}} \tag{.150}$$
$$+ \frac{-r_{ij}\frac{d}{dx_j}a_{ijk}\frac{d}{dy_j}r_{kj} + r_{ij}\frac{d}{dy_j}a_{ijk}\frac{d}{dx_j}r_{kj} - r_{kj}\frac{d}{dx_j}a_{ijk}\frac{d}{dy_j}r_{ij} + r_{kj}\frac{d}{dy_j}a_{ijk}\frac{d}{dx_j}r_{ij}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$
$$+ \frac{\left(a_{ijk}r_{ij}\frac{d}{dx_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dx_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dx_j}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dy_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}}$$
$$- \frac{2\left(a_{ijk}r_{ij}\frac{d}{dx_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dx_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dx_j}a_{ijk}\right)\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$
$$- \frac{2\left(a_{ijk}r_{ij}\frac{d}{dx_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dx_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dx_j}a_{ijk}\right)\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dz_j dx_j}\alpha_{ijk} = \frac{a_{ijk} r_{ij} \frac{d^2}{dz_j dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d^2}{dz_j dx_j} r_{ij} + a_{ijk} \frac{d}{dx_j} r_{ij} \frac{d}{dz_j} r_{kj} + a_{ijk} \frac{d}{dz_j} r_{ij} \frac{d}{dx_j} r_{kj} - r_{ij} r_{kj} \frac{d^2}{dz_j dx_j} a_{ijk}}{r_{ij}^2 r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$
$$+ \frac{-r_{ij} \frac{d}{dx_j} a_{ijk} \frac{d}{dz_j} r_{kj} + r_{ij} \frac{d}{dz_j} a_{ijk} \frac{d}{dx_j} r_{kj} - r_{kj} \frac{d}{dx_j} a_{ijk} \frac{d}{dz_j} r_{ij} + r_{kj} \frac{d}{dz_j} a_{ijk} \frac{d}{dx_j} r_{ij}}{r_{ij}^2 r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$
$$+ \frac{\left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right) \left(-\frac{a_{ijk}^2 \frac{d}{dz_j} r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2 \frac{d}{dz_j} r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk} \frac{d}{dz_j} a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}^2 \left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1\right)^{\frac{3}{2}}}$$
$$- \frac{2\left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right) \frac{d}{dz_j} r_{kj}}{r_{ij}^2 r_{kj}^3 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$
$$- \frac{2\left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right) \frac{d}{dz_j} r_{ij}}{r_{ij}^3 r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} \tag{.151}$$

$$\frac{d^2}{dx_k dx_j}\alpha_{ijk} = \frac{a_{ijk} r_{ij} \frac{d^2}{dx_k dx_j} r_{kj} + a_{ijk} \frac{d}{dx_j} r_{ij} \frac{d}{dx_k} r_{kj} - r_{ij} r_{kj} \frac{d^2}{dx_k dx_j} a_{ijk} - r_{ij} \frac{d}{dx_j} a_{ijk} \frac{d}{dx_k} r_{kj} + r_{ij} \frac{d}{dx_k} a_{ijk} \frac{d}{dx_j} r_{kj} + r_{kj} \frac{d}{dx_k} a_{ijk} \frac{d}{dx_j} r_{ij}}{r_{ij}^2 r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$
$$+ \frac{\left(-\frac{a_{ijk}^2 \frac{d}{dx_k} r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk} \frac{d}{dx_k} a_{ijk}}{r_{ij}^2 r_{kj}^2}\right) \left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right)}{r_{ij}^2 r_{kj}^2 \left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1\right)^{\frac{3}{2}}}$$
$$- \frac{2\left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right) \frac{d}{dx_k} r_{kj}}{r_{ij}^2 r_{kj}^3 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} \tag{.152}$$

$$\frac{d^2}{dy_k dx_j}\alpha_{ijk} = \frac{a_{ijk} r_{ij} \frac{d^2}{dy_k dx_j} r_{kj} + a_{ijk} \frac{d}{dx_j} r_{ij} \frac{d}{dy_k} r_{kj} - r_{ij} r_{kj} \frac{d^2}{dy_k dx_j} a_{ijk} - r_{ij} \frac{d}{dx_j} a_{ijk} \frac{d}{dy_k} r_{kj} + r_{ij} \frac{d}{dy_k} a_{ijk} \frac{d}{dx_j} r_{kj} + r_{kj} \frac{d}{dy_k} a_{ijk} \frac{d}{dx_j} r_{ij}}{r_{ij}^2 r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} \tag{.153}$$

$$+ \frac{\left(-\frac{a_{ijk}^2 \frac{d}{dy_k} r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk} \frac{d}{dy_k} a_{ijk}}{r_{ij}^2 r_{kj}^2}\right) \left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right)}{r_{ij}^2 r_{kj}^2 \left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1\right)^{\frac{3}{2}}}$$

$$- \frac{2 \left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right) \frac{d}{dy_k} r_{kj}}{r_{ij}^2 r_{kj}^3 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$

$$\frac{d^2}{dz_k dx_j}\alpha_{ijk} = \frac{a_{ijk} r_{ij} \frac{d^2}{dz_k dx_j} r_{kj} + a_{ijk} \frac{d}{dx_j} r_{ij} \frac{d}{dz_k} r_{kj} - r_{ij} r_{kj} \frac{d^2}{dz_k dx_j} a_{ijk} - r_{ij} \frac{d}{dx_j} a_{ijk} \frac{d}{dz_k} r_{kj} + r_{ij} \frac{d}{dz_k} a_{ijk} \frac{d}{dx_j} r_{kj} + r_{kj} \frac{d}{dz_k} a_{ijk} \frac{d}{dx_j} r_{ij}}{r_{ij}^2 r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} \tag{.154}$$

$$+ \frac{\left(-\frac{a_{ijk}^2 \frac{d}{dz_k} r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk} \frac{d}{dz_k} a_{ijk}}{r_{ij}^2 r_{kj}^2}\right) \left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right)}{r_{ij}^2 r_{kj}^2 \left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1\right)^{\frac{3}{2}}}$$

$$- \frac{2 \left(a_{ijk} r_{ij} \frac{d}{dx_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dx_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dx_j} a_{ijk}\right) \frac{d}{dz_k} r_{kj}}{r_{ij}^2 r_{kj}^3 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$

$$\frac{d^2}{dy_j^2}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dy_j^2}r_{kj} + a_{ijk}r_{kj}\frac{d^2}{dy_j^2}r_{ij} + 2a_{ijk}\frac{d}{dy_j}r_{ij}\frac{d}{dy_j}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dy_j^2}a_{ijk}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.155}$$
$$+ \frac{\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dy_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}}$$
$$- \frac{2\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\frac{d}{dy_j}r_{ij}}{r_{ij}^3 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dz_j dy_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dz_j dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d^2}{dz_j dy_j}r_{ij} + a_{ijk}\frac{d}{dy_j}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}\frac{d}{dz_j}r_{ij}\frac{d}{dy_j}r_{kj}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.156}$$
$$\frac{-r_{ij}r_{kj}\frac{d^2}{dz_j dy_j}a_{ijk} - r_{ij}\frac{d}{dy_j}a_{ijk}\frac{d}{dz_j}r_{kj} + r_{ij}\frac{d}{dz_j}a_{ijk}\frac{d}{dy_j}r_{kj} - r_{kj}\frac{d}{dy_j}a_{ijk}\frac{d}{dz_j}r_{ij} + r_{kj}\frac{d}{dz_j}a_{ijk}\frac{d}{dy_j}r_{ij}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$
$$+ \frac{\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dz_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}}$$
$$- \frac{2\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$
$$- \frac{2\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dx_k dy_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dy_j dx_k}r_{kj} + a_{ijk}\frac{d}{dy_j}r_{ij}\frac{d}{dx_k}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dy_j dx_k}a_{ijk} + r_{ij}\frac{d}{dx_k}a_{ijk}\frac{d}{dy_j}r_{kj} - r_{ij}\frac{d}{dy_j}a_{ijk}\frac{d}{dx_k}r_{kj} + r_{kj}\frac{d}{dx_k}a_{ijk}\frac{d}{dy_j}r_{ij}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.157}$$
$$+ \frac{\left(-\frac{a_{ijk}^2\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dx_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} - \frac{2\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dy_k dy_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dy_k dy_j}r_{kj} + a_{ijk}\frac{d}{dy_j}r_{ij}\frac{d}{dy_k}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dy_k dy_j}a_{ijk} - r_{ij}\frac{d}{dy_j}a_{ijk}\frac{d}{dy_k}r_{kj} + r_{ij}\frac{d}{dy_k}a_{ijk}\frac{d}{dy_j}r_{kj} + r_{kj}\frac{d}{dy_k}a_{ijk}\frac{d}{dy_j}r_{ij}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.158}$$
$$+ \frac{\left(-\frac{a_{ijk}^2\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dy_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} - \frac{2\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dz_k dy_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dz_k dy_j}r_{kj} + a_{ijk}\frac{d}{dy_j}r_{ij}\frac{d}{dz_k}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dz_k dy_j}a_{ijk} - r_{ij}\frac{d}{dy_j}a_{ijk}\frac{d}{dz_k}r_{kj} + r_{ij}\frac{d}{dz_k}a_{ijk}\frac{d}{dy_j}r_{kj} + r_{kj}\frac{d}{dz_k}a_{ijk}\frac{d}{dy_j}r_{ij}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.159}$$
$$+ \frac{\left(-\frac{a_{ijk}^2\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dz_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} - \frac{2\left(a_{ijk}r_{ij}\frac{d}{dy_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dy_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dy_j}a_{ijk}\right)\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dz_j^2}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dz_j^2}r_{kj} + a_{ijk}r_{kj}\frac{d^2}{dz_j^2}r_{ij} + 2a_{ijk}\frac{d}{dz_j}r_{ij}\frac{d}{dz_j}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dz_j^2}a_{ijk}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.160}$$
$$+ \frac{\left(a_{ijk}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dz_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dz_j}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^3} - \frac{a_{ijk}^2\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}^2} + \frac{a_{ijk}\frac{d}{dz_j}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}}$$
$$- \frac{2\left(a_{ijk}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dz_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dz_j}a_{ijk}\right)\frac{d}{dz_j}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dz_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dz_j}a_{ijk}\right)\frac{d}{dz_j}r_{ij}}{r_{ij}^3 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dx_k dz_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dz_j dx_k}r_{kj} + a_{ijk}\frac{d}{dz_j}r_{ij}\frac{d}{dx_k}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dz_j dx_k}a_{ijk} + r_{ij}\frac{d}{dx_k}a_{ijk}\frac{d}{dz_j}r_{kj} - r_{ij}\frac{d}{dz_j}a_{ijk}\frac{d}{dx_k}r_{kj} + r_{kj}\frac{d}{dx_k}a_{ijk}\frac{d}{dz_j}r_{ij}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.161}$$
$$+ \frac{\left(-\frac{a_{ijk}^2\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dx_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)\left(a_{ijk}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dz_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dz_j}a_{ijk}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} - \frac{2\left(a_{ijk}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dz_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dz_j}a_{ijk}\right)\frac{d}{dx_k}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dy_k dz_j}\alpha_{ijk} = \frac{a_{ijk}r_{ij}\frac{d^2}{dz_j dy_k}r_{kj} + a_{ijk}\frac{d}{dz_j}r_{ij}\frac{d}{dy_k}r_{kj} - r_{ij}r_{kj}\frac{d^2}{dz_j dy_k}a_{ijk} + r_{ij}\frac{d}{dy_k}a_{ijk}\frac{d}{dz_j}r_{kj} - r_{ij}\frac{d}{dz_j}a_{ijk}\frac{d}{dy_k}r_{kj} + r_{kj}\frac{d}{dy_k}a_{ijk}\frac{d}{dz_j}r_{ij}}{r_{ij}^2 r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.162}$$
$$+ \frac{\left(-\frac{a_{ijk}^2\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dy_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)\left(a_{ijk}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dz_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dz_j}a_{ijk}\right)}{r_{ij}^2 r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} - \frac{2\left(a_{ijk}r_{ij}\frac{d}{dz_j}r_{kj} + a_{ijk}r_{kj}\frac{d}{dz_j}r_{ij} - r_{ij}r_{kj}\frac{d}{dz_j}a_{ijk}\right)\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}$$

$$\frac{d^2}{dz_k dz_j}\alpha_{ijk} = \frac{a_{ijk} r_{ij} \frac{d^2}{dz_k dz_j} r_{kj} + a_{ijk} \frac{d}{dz_j} r_{ij} \frac{d}{dz_k} r_{kj} - r_{ij} r_{kj} \frac{d^2}{dz_k dz_j} a_{ijk} - r_{ij} \frac{d}{dz_j} a_{ijk} \frac{d}{dz_k} r_{kj} + r_{ij} \frac{d}{dz_k} a_{ijk} \frac{d}{dz_j} r_{kj} + r_{kj} \frac{d}{dz_k} a_{ijk} \frac{d}{dz_j} r_{ij}}{r_{ij}^2 r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} \tag{.163}$$
$$+ \frac{\left(-\frac{a_{ijk}^2 \frac{d}{dz_k} r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk} \frac{d}{dz_k} a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)\left(a_{ijk} r_{ij} \frac{d}{dz_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dz_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dz_j} a_{ijk}\right)}{r_{ij}^2 r_{kj}^2 \left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1\right)^{\frac{3}{2}}} - \frac{2\left(a_{ijk} r_{ij} \frac{d}{dz_j} r_{kj} + a_{ijk} r_{kj} \frac{d}{dz_j} r_{ij} - r_{ij} r_{kj} \frac{d}{dz_j} a_{ijk}\right) \frac{d}{dz_k} r_{kj}}{r_{ij}^2 r_{kj}^3 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$

$$\frac{d^2}{dx_k^2}\alpha_{ijk} = \frac{\left(a_{ijk} \frac{d}{dx_k} r_{kj} - r_{kj} \frac{d}{dx_k} a_{ijk}\right)\left(-\frac{a_{ijk}^2 \frac{d}{dx_k} r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk} \frac{d}{dx_k} a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij} r_{kj}^2 \left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1\right)^{\frac{3}{2}}} + \frac{a_{ijk} \frac{d^2}{dx_k^2} r_{kj} - r_{kj} \frac{d^2}{dx_k^2} a_{ijk}}{r_{ij} r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} - \frac{2\left(a_{ijk} \frac{d}{dx_k} r_{kj} - r_{kj} \frac{d}{dx_k} a_{ijk}\right) \frac{d}{dx_k} r_{kj}}{r_{ij} r_{kj}^3 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} \tag{.164}$$

$$\frac{d^2}{dy_k dx_k}\alpha_{ijk} = \frac{\left(a_{ijk} \frac{d}{dx_k} r_{kj} - r_{kj} \frac{d}{dx_k} a_{ijk}\right)\left(-\frac{a_{ijk}^2 \frac{d}{dy_k} r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk} \frac{d}{dy_k} a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij} r_{kj}^2 \left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1\right)^{\frac{3}{2}}} \tag{.165}$$
$$+ \frac{a_{ijk} \frac{d^2}{dy_k dx_k} r_{kj} - r_{kj} \frac{d^2}{dy_k dx_k} a_{ijk} - \frac{d}{dx_k} a_{ijk} \frac{d}{dy_k} r_{kj} + \frac{d}{dy_k} a_{ijk} \frac{d}{dx_k} r_{kj}}{r_{ij} r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} - \frac{2\left(a_{ijk} \frac{d}{dx_k} r_{kj} - r_{kj} \frac{d}{dx_k} a_{ijk}\right) \frac{d}{dy_k} r_{kj}}{r_{ij} r_{kj}^3 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$

$$\frac{d^2}{dz_k dx_k}\alpha_{ijk} = \frac{\left(a_{ijk} \frac{d}{dx_k} r_{kj} - r_{kj} \frac{d}{dx_k} a_{ijk}\right)\left(-\frac{a_{ijk}^2 \frac{d}{dz_k} r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk} \frac{d}{dz_k} a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij} r_{kj}^2 \left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1\right)^{\frac{3}{2}}} \tag{.166}$$
$$+ \frac{a_{ijk} \frac{d^2}{dz_k dx_k} r_{kj} - r_{kj} \frac{d^2}{dz_k dx_k} a_{ijk} - \frac{d}{dx_k} a_{ijk} \frac{d}{dz_k} r_{kj} + \frac{d}{dz_k} a_{ijk} \frac{d}{dx_k} r_{kj}}{r_{ij} r_{kj}^2 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}} - \frac{2\left(a_{ijk} \frac{d}{dx_k} r_{kj} - r_{kj} \frac{d}{dx_k} a_{ijk}\right) \frac{d}{dz_k} r_{kj}}{r_{ij} r_{kj}^3 \sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2} + 1}}$$

$$\frac{d^2}{dy_k^2}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dy_k}r_{kj} - r_{kj}\frac{d}{dy_k}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dy_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dy_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dy_k^2}r_{kj} - r_{kj}\frac{d^2}{dy_k^2}a_{ijk}}{r_{ij}r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dy_k}r_{kj} - r_{kj}\frac{d}{dy_k}a_{ijk}\right)\frac{d}{dy_k}r_{kj}}{r_{ij}r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.167}$$

$$\begin{aligned}\frac{d^2}{dz_k dy_k}\alpha_{ijk} = &\frac{\left(a_{ijk}\frac{d}{dy_k}r_{kj} - r_{kj}\frac{d}{dy_k}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dz_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} \\ &+ \frac{a_{ijk}\frac{d^2}{dz_k dy_k}r_{kj} - r_{kj}\frac{d^2}{dz_k dy_k}a_{ijk} - \frac{d}{dy_k}a_{ijk}\frac{d}{dz_k}r_{kj} + \frac{d}{dz_k}a_{ijk}\frac{d}{dy_k}r_{kj}}{r_{ij}r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dy_k}r_{kj} - r_{kj}\frac{d}{dy_k}a_{ijk}\right)\frac{d}{dz_k}r_{kj}}{r_{ij}r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}}\end{aligned} \tag{.168}$$

$$\frac{d^2}{dz_k^2}\alpha_{ijk} = \frac{\left(a_{ijk}\frac{d}{dz_k}r_{kj} - r_{kj}\frac{d}{dz_k}a_{ijk}\right)\left(-\frac{a_{ijk}^2\frac{d}{dz_k}r_{kj}}{r_{ij}^2 r_{kj}^3} + \frac{a_{ijk}\frac{d}{dz_k}a_{ijk}}{r_{ij}^2 r_{kj}^2}\right)}{r_{ij}r_{kj}^2\left(-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1\right)^{\frac{3}{2}}} + \frac{a_{ijk}\frac{d^2}{dz_k^2}r_{kj} - r_{kj}\frac{d^2}{dz_k^2}a_{ijk}}{r_{ij}r_{kj}^2\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} - \frac{2\left(a_{ijk}\frac{d}{dz_k}r_{kj} - r_{kj}\frac{d}{dz_k}a_{ijk}\right)\frac{d}{dz_k}r_{kj}}{r_{ij}r_{kj}^3\sqrt{-\frac{a_{ijk}^2}{r_{ij}^2 r_{kj}^2}+1}} \tag{.169}$$

The derivatives of a_{ijk} with respect to the coordinates x, y and z of the participating atoms i, j and k are defined in Section B. The derivatives of r_{ij} and r_{kj} with respect to the coordinates x, y and z of the participating atoms i, j and k are defined in Section B. The derivatives of r_{ij} and r_{kj} are identical, only differing the atomic indices i and k.

Derivatives of a_{ijk}

The first and second order derivatives of a_{ijk} with respect to the coordinates x, y and z of the participating atoms i, j and k are:

$$\frac{d}{dx_i}a_{ijk} = x_{kj} \tag{.170}$$

$$\frac{d}{dy_i}a_{ijk} = y_{kj} \tag{.171}$$

$$\frac{d}{dz_i}a_{ijk} = z_{kj} \tag{.172}$$

$$\frac{d}{dx_j}a_{ijk} = -x_i + 2x_j - x_k \tag{.173}$$

$$\frac{d}{dy_j}a_{ijk} = -y_i + 2y_j - y_k \tag{.174}$$

$$\frac{d}{dz_j}a_{ijk} = -z_i + 2z_j - z_k \tag{.175}$$

$$\frac{d}{dx_k}a_{ijk} = -x_{ji} \tag{.176}$$

$$\frac{d}{dy_k}a_{ijk} = -y_{ji} \tag{.177}$$

$$\frac{d}{dz_k}a_{ijk} = -z_{ji} \tag{.178}$$

$$\frac{d^2}{dx_i^2}a_{ijk} = 0 \tag{.179}$$

$$\frac{d^2}{dy_i dx_i}a_{ijk} = 0 \tag{.180}$$

$$\frac{d^2}{dz_i dx_i}a_{ijk} = 0 \tag{.181}$$

$$\frac{d^2}{dx_j dx_i}a_{ijk} = -1 \tag{.182}$$

$$\frac{d^2}{dy_j dx_i}a_{ijk} = 0 \tag{.183}$$

$$\frac{d^2}{dz_j dx_i} a_{ijk} = 0 \tag{.184}$$

$$\frac{d^2}{dx_k dx_i} a_{ijk} = 1 \tag{.185}$$

$$\frac{d^2}{dy_k dx_i} a_{ijk} = 0 \tag{.186}$$

$$\frac{d^2}{dz_k dx_i} a_{ijk} = 0 \tag{.187}$$

$$\frac{d^2}{dy_i^2} a_{ijk} = 0 \tag{.188}$$

$$\frac{d^2}{dz_i dy_i} a_{ijk} = 0 \tag{.189}$$

$$\frac{d^2}{dx_j dy_i} a_{ijk} = 0 \tag{.190}$$

$$\frac{d^2}{dy_j dy_i} a_{ijk} = -1 \tag{.191}$$

$$\frac{d^2}{dz_j dy_i} a_{ijk} = 0 \tag{.192}$$

$$\frac{d^2}{dx_k dy_i} a_{ijk} = 0 \tag{.193}$$

$$\frac{d^2}{dy_k dy_i} a_{ijk} = 1 \tag{.194}$$

$$\frac{d^2}{dz_k dy_i} a_{ijk} = 0 \tag{.195}$$

$$\frac{d^2}{dz_i^2} a_{ijk} = 0 \tag{.196}$$

$$\frac{d^2}{dx_j dz_i} a_{ijk} = 0 \tag{.197}$$

$$\frac{d^2}{dy_j dz_i} a_{ijk} = 0 \tag{.198}$$

$$\frac{d^2}{dz_j dz_i} a_{ijk} = -1 \tag{.199}$$

$$\frac{d^2}{dx_k dz_i} a_{ijk} = 0 \tag{.200}$$

$$\frac{d^2}{dy_k dz_i} a_{ijk} = 0 \tag{.201}$$

$$\frac{d^2}{dz_k dz_i} a_{ijk} = 1 \tag{.202}$$

$$\frac{d^2}{dx_j^2} a_{ijk} = 2 \tag{.203}$$

$$\frac{d^2}{dy_j dx_j} a_{ijk} = 0 \tag{.204}$$

$$\frac{d^2}{dz_j dx_j} a_{ijk} = 0 \tag{.205}$$

$$\frac{d^2}{dx_k dx_j} a_{ijk} = -1 \tag{.206}$$

$$\frac{d^2}{dy_k dx_j} a_{ijk} = 0 \tag{.207}$$

$$\frac{d^2}{dz_k dx_j} a_{ijk} = 0 \tag{.208}$$

$$\frac{d^2}{dy_j^2} a_{ijk} = 2 \tag{.209}$$

$$\frac{d^2}{dz_j dy_j} a_{ijk} = 0 \tag{.210}$$

$$\frac{d^2}{dx_k dy_j} a_{ijk} = 0 \tag{.211}$$

$$\frac{d^2}{dy_k dy_j} a_{ijk} = -1 \tag{.212}$$

$$\frac{d^2}{dz_k dy_j} a_{ijk} = 0 \tag{.213}$$

$$\frac{d^2}{dz_j^2} a_{ijk} = 2 \tag{.214}$$

$$\frac{d^2}{dx_k dz_j} a_{ijk} = 0 \tag{.215}$$

$$\frac{d^2}{dy_k dz_j} a_{ijk} = 0 \tag{.216}$$

$$\frac{d^2}{dz_k dz_j} a_{ijk} = -1 \tag{.217}$$

$$\frac{d^2}{dx_k^2} a_{ijk} = 0 \tag{.218}$$

$$\frac{d^2}{dy_k dx_k} a_{ijk} = 0 \tag{.219}$$

$$\frac{d^2}{dz_k dx_k} a_{ijk} = 0 \tag{.220}$$

$$\frac{d^2}{dy_k^2} a_{ijk} = 0 \tag{.221}$$

$$\frac{d^2}{dz_k dy_k} a_{ijk} = 0 \tag{.222}$$

$$\frac{d^2}{dz_k^2} a_{ijk} = 0 \tag{.223}$$

Dihedral Potential

The dihedral harmonic potential between three atoms i, j, k and l is given by

$$E_{\text{dihedral},ijkl} = \sum k_{\text{dihedral}} \left(\left(\sin\theta_{ijkl}^0 - \sin\theta_{ijkl} \right)^2 + \left(\cos\theta_{ijkl}^0 - \cos\theta_{ijkl} \right)^2 \right) + \sum \frac{Q}{r_{ij}}$$

The cosine of the dihedral angle between four atoms i, j, k and l is given by

$$\cos\theta_{ijkl} = \frac{\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right)\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right) + \left(x_{ji}z_{kj} - x_{kj}z_{ji}\right)\left(x_{kj}z_{lk} - x_{lk}z_{kj}\right) + \left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right)}{\sqrt{\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right)^2 + \left(x_{ji}z_{kj} - x_{kj}z_{ji}\right)^2 + \left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)^2}\sqrt{\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)^2 + \left(x_{kj}z_{lk} - x_{lk}z_{kj}\right)^2 + \left(y_{kj}z_{lk} - y_{lk}z_{kj}\right)^2}} \tag{.224}$$

$$= \frac{a_{ijkl}}{b_{ijk}b_{jkl}} \,, \tag{.225}$$

where a_{ijkl} is the dot product of the cross product of $\mathbf{r}_{ji}$ and $\mathbf{r}_{kj}$ and the cross product of $\mathbf{r}_{kj}$ and $\mathbf{r}_{lk}$. b_{ijk} is the norm of the first cross product and b_{jkl} is the norm of the second cross product. The sine of the dihedral angle between four atoms i, j, k and l is given by

$$\sin\theta_{ijkl} = \frac{\sqrt{x_{kj}^2 + y_{kj}^2 + z_{kj}^2}\left(x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) - y_{ji}\left(x_{kj}z_{lk} - x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)\right)}{\sqrt{\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right)^2 + \left(x_{ji}z_{kj} - x_{kj}z_{ji}\right)^2 + \left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)^2}\sqrt{\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)^2 + \left(x_{kj}z_{lk} - x_{lk}z_{kj}\right)^2 + \left(y_{kj}z_{lk} - y_{lk}z_{kj}\right)^2}} \tag{.226}$$

$$= \frac{d_{ijkl}}{b_{ijk}b_{jkl}} \,, \tag{.227}$$

where d_{ijkl} is the dot product of $\mathbf{r}_{ji}$ and the cross product of $\mathbf{r}_{kj}$ and $\mathbf{r}_{lk}$ multiplied by the norm of $\mathbf{r}_{kj}$.

Derivatives of $E_{\textbf{dihedral},ijkl}$

The first and second order derivatives of $E_{\text{dihedral},ijkl}$ with respect to the coordinates x, y and z of the participating atoms i, j, k and l are:

$$\frac{d}{dx_i}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\left(-\left(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\right)\frac{d}{dx_i}\cos\theta_{ijkl} - \left(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\right)\frac{d}{dx_i}\sin\theta_{ijkl}\right) \tag{.228}$$

$$\frac{d}{dy_i}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\left(-\left(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\right)\frac{d}{dy_i}\cos\theta_{ijkl} - \left(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\right)\frac{d}{dy_i}\sin\theta_{ijkl}\right) \tag{.229}$$

$$\frac{d}{dz_i}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\left(-\left(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\right)\frac{d}{dz_i}\cos\theta_{ijkl} - \left(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\right)\frac{d}{dz_i}\sin\theta_{ijkl}\right) \tag{.230}$$

$$\frac{d}{dx_j}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\left(-\left(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\right)\frac{d}{dx_j}\cos\theta_{ijkl} - \left(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\right)\frac{d}{dx_j}\sin\theta_{ijkl}\right) \tag{.231}$$

$$\frac{d}{dy_j}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\left(-\left(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\right)\frac{d}{dy_j}\cos\theta_{ijkl} - \left(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\right)\frac{d}{dy_j}\sin\theta_{ijkl}\right) \tag{.232}$$

$$\frac{d}{dz_j}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\left(-\left(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\right)\frac{d}{dz_j}\cos\theta_{ijkl} - \left(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\right)\frac{d}{dz_j}\sin\theta_{ijkl}\right) \tag{.233}$$

$$\frac{d}{dx_k}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\left(-\left(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\right)\frac{d}{dx_k}\cos\theta_{ijkl} - \left(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\right)\frac{d}{dx_k}\sin\theta_{ijkl}\right) \tag{.234}$$

$$\frac{d}{dy_k}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d}{dy_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d}{dy_k}\sin\theta_{ijkl}\big) \tag{.235}$$

$$\frac{d}{dz_k}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d}{dz_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.236}$$

$$\frac{d}{dx_l}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d}{dx_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d}{dx_l}\sin\theta_{ijkl}\big) \tag{.237}$$

$$\frac{d}{dy_l}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d}{dy_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d}{dy_l}\sin\theta_{ijkl}\big) \tag{.238}$$

$$\frac{d}{dz_l}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d}{dz_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.239}$$

$$\frac{d^2}{dx_i^2}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dx_i^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dx_i^2}\sin\theta_{ijkl} + \big(\frac{d}{dx_i}\cos\theta_{ijkl}\big)^2 + \big(\frac{d}{dx_i}\sin\theta_{ijkl}\big)^2\big) \tag{.240}$$

$$\begin{aligned}\frac{d^2}{dy_i dx_i}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(&-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dy_i dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dy_i dx_i}\sin\theta_{ijkl}\\ &+ \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dy_i}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dy_i}\sin\theta_{ijkl}\big)\end{aligned} \tag{.241}$$

$$\frac{d^2}{dz_i dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_i dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_i dx_i}\sin\theta_{ijkl} + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dz_i}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dz_i}\sin\theta_{ijkl}\big) \tag{.242}$$

$$\frac{d^2}{dx_j dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dx_j dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dx_j dx_i}\sin\theta_{ijkl} + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dx_j}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dx_j}\sin\theta_{ijkl}\big) \tag{.243}$$

$$\frac{d^2}{dy_j dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_j dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_j dx_i}\sin\theta_{ijkl} + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dy_j}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dy_j}\sin\theta_{ijkl}\big) \tag{.244}$$

$$\frac{d^2}{dz_j dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_j dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_j dx_i}\sin\theta_{ijkl} + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dz_j}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dz_j}\sin\theta_{ijkl}\big) \tag{.245}$$

$$\frac{d^2}{dx_k dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dx_k dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dx_k dx_i}\sin\theta_{ijkl} + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dx_k}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dx_k}\sin\theta_{ijkl}\big) \tag{.246}$$

$$\frac{d^2}{dy_k dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_k dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_k dx_i}\sin\theta_{ijkl} \\ + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dy_k}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dy_k}\sin\theta_{ijkl}\big) \tag{.247}$$

$$\frac{d^2}{dz_k dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_k dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_k dx_i}\sin\theta_{ijkl} \\ + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.248}$$

$$\frac{d^2}{dx_l dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dx_l dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dx_l dx_i}\sin\theta_{ijkl} \\ + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dx_l}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dx_l}\sin\theta_{ijkl}\big) \tag{.249}$$

$$\frac{d^2}{dy_l dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_l dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_l dx_i}\sin\theta_{ijkl} \\ + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dy_l}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dy_l}\sin\theta_{ijkl}\big) \tag{.250}$$

$$\frac{d^2}{dz_l dx_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l dx_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l dx_i}\sin\theta_{ijkl} \\ + \frac{d}{dx_i}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dx_i}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.251}$$

$$\frac{d^2}{dy_i^2}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dy_i^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dy_i^2}\sin\theta_{ijkl} \\ + (\frac{d}{dy_i}\cos\theta_{ijkl})^2 + (\frac{d}{dy_i}\sin\theta_{ijkl})^2\big) \tag{.252}$$

$$\frac{d^2}{dz_i dy_i}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dz_i dy_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dz_i dy_i}\sin\theta_{ijkl} \\ + \frac{d}{dy_i}\cos\theta_{ijkl}\frac{d}{dz_i}\cos\theta_{ijkl} + \frac{d}{dy_i}\sin\theta_{ijkl}\frac{d}{dz_i}\sin\theta_{ijkl}\big) \tag{.253}$$

$$\frac{d^2}{dx_j dy_i}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dy_i dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dy_i dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dy_i}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dy_i}\sin\theta_{ijkl}\big) \tag{.254}$$

$$\frac{d^2}{dy_j dy_i}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dy_j dy_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dy_j dy_i}\sin\theta_{ijkl} \\ + \frac{d}{dy_i}\cos\theta_{ijkl}\frac{d}{dy_j}\cos\theta_{ijkl} + \frac{d}{dy_i}\sin\theta_{ijkl}\frac{d}{dy_j}\sin\theta_{ijkl}\big) \tag{.255}$$

$$\frac{d^2}{dz_j dy_i}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dz_j dy_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dz_j dy_i}\sin\theta_{ijkl} \\ + \frac{d}{dy_i}\cos\theta_{ijkl}\frac{d}{dz_j}\cos\theta_{ijkl} + \frac{d}{dy_i}\sin\theta_{ijkl}\frac{d}{dz_j}\sin\theta_{ijkl}\big) \tag{.256}$$

$$\frac{d^2}{dx_k dy_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_i dx_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_i dx_k}\sin\theta_{ijkl} \\ + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dy_i}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dy_i}\sin\theta_{ijkl}\big) \tag{.257}$$

$$\frac{d^2}{dy_k dy_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_k dy_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_k dy_i}\sin\theta_{ijkl} \\ + \frac{d}{dy_i}\cos\theta_{ijkl}\frac{d}{dy_k}\cos\theta_{ijkl} + \frac{d}{dy_i}\sin\theta_{ijkl}\frac{d}{dy_k}\sin\theta_{ijkl}\big) \tag{.258}$$

$$\frac{d^2}{dz_k dy_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_k dy_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_k dy_i}\sin\theta_{ijkl} \\ + \frac{d}{dy_i}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dy_i}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.259}$$

$$\frac{d^2}{dx_l dy_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_i dx_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_i dx_l}\sin\theta_{ijkl} \\ + \frac{d}{dx_l}\cos\theta_{ijkl}\frac{d}{dy_i}\cos\theta_{ijkl} + \frac{d}{dx_l}\sin\theta_{ijkl}\frac{d}{dy_i}\sin\theta_{ijkl}\big) \tag{.260}$$

$$\frac{d^2}{dy_l dy_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_l dy_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_l dy_i}\sin\theta_{ijkl} \\ + \frac{d}{dy_i}\cos\theta_{ijkl}\frac{d}{dy_l}\cos\theta_{ijkl} + \frac{d}{dy_i}\sin\theta_{ijkl}\frac{d}{dy_l}\sin\theta_{ijkl}\big) \tag{.261}$$

$$\frac{d^2}{dz_l dy_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l dy_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l dy_i}\sin\theta_{ijkl} \\ + \frac{d}{dy_i}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dy_i}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.262}$$

$$\frac{d^2}{dz_i^2} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_i^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_i^2}\sin\theta_{ijkl} \\ + (\frac{d}{dz_i}\cos\theta_{ijkl})^2 + (\frac{d}{dz_i}\sin\theta_{ijkl})^2\big) \tag{.263}$$

$$\frac{d^2}{dx_j dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_i dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_i dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dz_i}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dz_i}\sin\theta_{ijkl}\big) \tag{.264}$$

$$\frac{d^2}{dy_j dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_i dy_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_i dy_j}\sin\theta_{ijkl} \\ + \frac{d}{dy_j}\cos\theta_{ijkl}\frac{d}{dz_i}\cos\theta_{ijkl} + \frac{d}{dy_j}\sin\theta_{ijkl}\frac{d}{dz_i}\sin\theta_{ijkl}\big) \tag{.265}$$

$$\frac{d^2}{dz_j dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_j dz_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_j dz_i}\sin\theta_{ijkl} \\ + \frac{d}{dz_i}\cos\theta_{ijkl}\frac{d}{dz_j}\cos\theta_{ijkl} + \frac{d}{dz_i}\sin\theta_{ijkl}\frac{d}{dz_j}\sin\theta_{ijkl}\big) \tag{.266}$$

$$\frac{d^2}{dx_k dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_i dx_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_i dx_k}\sin\theta_{ijkl} \\ + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dz_i}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dz_i}\sin\theta_{ijkl}\big) \tag{.267}$$

$$\frac{d^2}{dy_k dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_i dy_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_i dy_k}\sin\theta_{ijkl} \\ + \frac{d}{dy_k}\cos\theta_{ijkl}\frac{d}{dz_i}\cos\theta_{ijkl} + \frac{d}{dy_k}\sin\theta_{ijkl}\frac{d}{dz_i}\sin\theta_{ijkl}\big) \tag{.268}$$

$$\frac{d^2}{dz_k dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_k dz_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_k dz_i}\sin\theta_{ijkl} \\ + \frac{d}{dz_i}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dz_i}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.269}$$

$$\frac{d^2}{dx_l dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_i dx_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_i dx_l}\sin\theta_{ijkl} \\ + \frac{d}{dx_l}\cos\theta_{ijkl}\frac{d}{dz_i}\cos\theta_{ijkl} + \frac{d}{dx_l}\sin\theta_{ijkl}\frac{d}{dz_i}\sin\theta_{ijkl}\big) \tag{.270}$$

$$\frac{d^2}{dy_l dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_i dy_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_i dy_l}\sin\theta_{ijkl} \\ + \frac{d}{dy_l}\cos\theta_{ijkl}\frac{d}{dz_i}\cos\theta_{ijkl} + \frac{d}{dy_l}\sin\theta_{ijkl}\frac{d}{dz_i}\sin\theta_{ijkl}\big) \tag{.271}$$

$$\frac{d^2}{dz_l dz_i} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l dz_i}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l dz_i}\sin\theta_{ijkl} \\ + \frac{d}{dz_i}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dz_i}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.272}$$

$$\frac{d^2}{dx_j^2} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dx_j^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dx_j^2}\sin\theta_{ijkl} \\ + (\frac{d}{dx_j}\cos\theta_{ijkl})^2 + (\frac{d}{dx_j}\sin\theta_{ijkl})^2\big) \tag{.273}$$

$$\frac{d^2}{dy_j dx_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_j dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_j dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dy_j}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dy_j}\sin\theta_{ijkl}\big) \tag{.274}$$

$$\frac{d^2}{dz_j dx_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_j dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_j dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dz_j}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dz_j}\sin\theta_{ijkl}\big) \tag{.275}$$

$$\frac{d^2}{dx_k dx_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dx_k dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dx_k dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dx_k}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dx_k}\sin\theta_{ijkl}\big) \tag{.276}$$

$$\frac{d^2}{dy_k dx_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_k dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_k dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dy_k}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dy_k}\sin\theta_{ijkl}\big) \tag{.277}$$

$$\frac{d^2}{dz_k dx_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_k dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_k dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.278}$$

$$\frac{d^2}{dx_l dx_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dx_l dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dx_l dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dx_l}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dx_l}\sin\theta_{ijkl}\big) \tag{.279}$$

$$\frac{d^2}{dy_l dx_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_l dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_l dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dy_l}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dy_l}\sin\theta_{ijkl}\big) \tag{.280}$$

$$\frac{d^2}{dz_l dx_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l dx_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l dx_j}\sin\theta_{ijkl} \\ + \frac{d}{dx_j}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dx_j}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.281}$$

$$\frac{d^2}{dy_j^2}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dy_j^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dy_j^2}\sin\theta_{ijkl} \\ + (\frac{d}{dy_j}\cos\theta_{ijkl})^2 + (\frac{d}{dy_j}\sin\theta_{ijkl})^2\big) \tag{.282}$$

$$\frac{d^2}{dz_j dy_j}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dz_j dy_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dz_j dy_j}\sin\theta_{ijkl} \\ + \frac{d}{dy_j}\cos\theta_{ijkl}\frac{d}{dz_j}\cos\theta_{ijkl} + \frac{d}{dy_j}\sin\theta_{ijkl}\frac{d}{dz_j}\sin\theta_{ijkl}\big) \tag{.283}$$

$$\frac{d^2}{dx_k dy_j}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dy_j dx_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dy_j dx_k}\sin\theta_{ijkl} \\ + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dy_j}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dy_j}\sin\theta_{ijkl}\big) \tag{.284}$$

$$\frac{d^2}{dy_k dy_j}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dy_k dy_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dy_k dy_j}\sin\theta_{ijkl} \\ + \frac{d}{dy_j}\cos\theta_{ijkl}\frac{d}{dy_k}\cos\theta_{ijkl} + \frac{d}{dy_j}\sin\theta_{ijkl}\frac{d}{dy_k}\sin\theta_{ijkl}\big) \tag{.285}$$

$$\frac{d^2}{dz_k dy_j}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl}+\cos\theta^0_{ijk})\frac{d^2}{dz_k dy_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl}+\sin\theta^0_{ijk})\frac{d^2}{dz_k dy_j}\sin\theta_{ijkl} \\ + \frac{d}{dy_j}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dy_j}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.286}$$

$$\frac{d^2}{dx_l dy_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_j dx_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_j dx_l}\sin\theta_{ijkl} \\ + \frac{d}{dx_l}\cos\theta_{ijkl}\frac{d}{dy_j}\cos\theta_{ijkl} + \frac{d}{dx_l}\sin\theta_{ijkl}\frac{d}{dy_j}\sin\theta_{ijkl}\big) \tag{.287}$$

$$\frac{d^2}{dy_l dy_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_l dy_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_l dy_j}\sin\theta_{ijkl} \\ + \frac{d}{dy_j}\cos\theta_{ijkl}\frac{d}{dy_l}\cos\theta_{ijkl} + \frac{d}{dy_j}\sin\theta_{ijkl}\frac{d}{dy_l}\sin\theta_{ijkl}\big) \tag{.288}$$

$$\frac{d^2}{dz_l dy_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l dy_j}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l dy_j}\sin\theta_{ijkl} \\ + \frac{d}{dy_j}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dy_j}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.289}$$

$$\frac{d^2}{dz_j^2} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_j^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_j^2}\sin\theta_{ijkl} + (\frac{d}{dz_j}\cos\theta_{ijkl})^2 \\ + (\frac{d}{dz_j}\sin\theta_{ijkl})^2\big) \tag{.290}$$

$$\frac{d^2}{dx_k dz_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_j dx_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_j dx_k}\sin\theta_{ijkl} \\ + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dz_j}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dz_j}\sin\theta_{ijkl}\big) \tag{.291}$$

$$\frac{d^2}{dy_k dz_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}} \big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}) \frac{d^2}{dz_j dy_k} \cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk}) \frac{d^2}{dz_j dy_k} \sin\theta_{ijkl} \\ + \frac{d}{dy_k} \cos\theta_{ijkl} \frac{d}{dz_j} \cos\theta_{ijkl} + \frac{d}{dy_k} \sin\theta_{ijkl} \frac{d}{dz_j} \sin\theta_{ijkl} \big) \tag{.292}$$

$$\frac{d^2}{dz_k dz_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}} \big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}) \frac{d^2}{dz_k dz_j} \cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk}) \frac{d^2}{dz_k dz_j} \sin\theta_{ijkl} \\ + \frac{d}{dz_j} \cos\theta_{ijkl} \frac{d}{dz_k} \cos\theta_{ijkl} + \frac{d}{dz_j} \sin\theta_{ijkl} \frac{d}{dz_k} \sin\theta_{ijkl} \big) \tag{.293}$$

$$\frac{d^2}{dx_l dz_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}} \big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}) \frac{d^2}{dz_j dx_l} \cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk}) \frac{d^2}{dz_j dx_l} \sin\theta_{ijkl} \\ + \frac{d}{dx_l} \cos\theta_{ijkl} \frac{d}{dz_j} \cos\theta_{ijkl} + \frac{d}{dx_l} \sin\theta_{ijkl} \frac{d}{dz_j} \sin\theta_{ijkl} \big) \tag{.294}$$

$$\frac{d^2}{dy_l dz_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}} \big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}) \frac{d^2}{dz_j dy_l} \cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk}) \frac{d^2}{dz_j dy_l} \sin\theta_{ijkl} \\ + \frac{d}{dy_l} \cos\theta_{ijkl} \frac{d}{dz_j} \cos\theta_{ijkl} + \frac{d}{dy_l} \sin\theta_{ijkl} \frac{d}{dz_j} \sin\theta_{ijkl} \big) \tag{.295}$$

$$\frac{d^2}{dz_l dz_j} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}} \big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}) \frac{d^2}{dz_l dz_j} \cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk}) \frac{d^2}{dz_l dz_j} \sin\theta_{ijkl} \\ + \frac{d}{dz_j} \cos\theta_{ijkl} \frac{d}{dz_l} \cos\theta_{ijkl} + \frac{d}{dz_j} \sin\theta_{ijkl} \frac{d}{dz_l} \sin\theta_{ijkl} \big) \tag{.296}$$

$$\frac{d^2}{dx_k^2}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl}+\cos\theta_{ijk}^0\big)\frac{d^2}{dx_k^2}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl}+\sin\theta_{ijk}^0\big)\frac{d^2}{dx_k^2}\sin\theta_{ijkl} + \big(\frac{d}{dx_k}\cos\theta_{ijkl}\big)^2 + \big(\frac{d}{dx_k}\sin\theta_{ijkl}\big)^2\big) \tag{.297}$$

$$\frac{d^2}{dy_k dx_k}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl}+\cos\theta_{ijk}^0\big)\frac{d^2}{dy_k dx_k}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl}+\sin\theta_{ijk}^0\big)\frac{d^2}{dy_k dx_k}\sin\theta_{ijkl} + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dy_k}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dy_k}\sin\theta_{ijkl}\big) \tag{.298}$$

$$\frac{d^2}{dz_k dx_k}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl}+\cos\theta_{ijk}^0\big)\frac{d^2}{dz_k dx_k}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl}+\sin\theta_{ijk}^0\big)\frac{d^2}{dz_k dx_k}\sin\theta_{ijkl} + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.299}$$

$$\frac{d^2}{dx_l dx_k}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl}+\cos\theta_{ijk}^0\big)\frac{d^2}{dx_l dx_k}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl}+\sin\theta_{ijk}^0\big)\frac{d^2}{dx_l dx_k}\sin\theta_{ijkl} + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dx_l}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dx_l}\sin\theta_{ijkl}\big) \tag{.300}$$

$$\frac{d^2}{dy_l dx_k}E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl}+\cos\theta_{ijk}^0\big)\frac{d^2}{dy_l dx_k}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl}+\sin\theta_{ijk}^0\big)\frac{d^2}{dy_l dx_k}\sin\theta_{ijkl} + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dy_l}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dy_l}\sin\theta_{ijkl}\big) \tag{.301}$$

$$\frac{d^2}{dz_l dx_k} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l dx_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l dx_k}\sin\theta_{ijkl} \\ + \frac{d}{dx_k}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dx_k}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.302}$$

$$\frac{d^2}{dy_k^2} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_k^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_k^2}\sin\theta_{ijkl} + (\frac{d}{dy_k}\cos\theta_{ijkl})^2 \\ + (\frac{d}{dy_k}\sin\theta_{ijkl})^2\big) \tag{.303}$$

$$\frac{d^2}{dz_k dy_k} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_k dy_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_k dy_k}\sin\theta_{ijkl} \\ + \frac{d}{dy_k}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dy_k}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.304}$$

$$\frac{d^2}{dx_l dy_k} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_k dx_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_k dx_l}\sin\theta_{ijkl} \\ + \frac{d}{dx_l}\cos\theta_{ijkl}\frac{d}{dy_k}\cos\theta_{ijkl} + \frac{d}{dx_l}\sin\theta_{ijkl}\frac{d}{dy_k}\sin\theta_{ijkl}\big) \tag{.305}$$

$$\frac{d^2}{dy_l dy_k} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_l dy_k}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_l dy_k}\sin\theta_{ijkl} \\ + \frac{d}{dy_k}\cos\theta_{ijkl}\frac{d}{dy_l}\cos\theta_{ijkl} + \frac{d}{dy_k}\sin\theta_{ijkl}\frac{d}{dy_l}\sin\theta_{ijkl}\big) \tag{.306}$$

$$\frac{d^2}{dz_l dy_k} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\big)\frac{d^2}{dz_l dy_k}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\big)\frac{d^2}{dz_l dy_k}\sin\theta_{ijkl} + \frac{d}{dy_k}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dy_k}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.307}$$

$$\frac{d^2}{dz_k^2} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\big)\frac{d^2}{dz_k^2}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\big)\frac{d^2}{dz_k^2}\sin\theta_{ijkl} + \big(\frac{d}{dz_k}\cos\theta_{ijkl}\big)^2 + \big(\frac{d}{dz_k}\sin\theta_{ijkl}\big)^2\big) \tag{.308}$$

$$\frac{d^2}{dx_l dz_k} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\big)\frac{d^2}{dz_k dx_l}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\big)\frac{d^2}{dz_k dx_l}\sin\theta_{ijkl} + \frac{d}{dx_l}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dx_l}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.309}$$

$$\frac{d^2}{dy_l dz_k} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\big)\frac{d^2}{dz_k dy_l}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\big)\frac{d^2}{dz_k dy_l}\sin\theta_{ijkl} + \frac{d}{dy_l}\cos\theta_{ijkl}\frac{d}{dz_k}\cos\theta_{ijkl} + \frac{d}{dy_l}\sin\theta_{ijkl}\frac{d}{dz_k}\sin\theta_{ijkl}\big) \tag{.310}$$

$$\frac{d^2}{dz_l dz_k} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\big)\frac{d^2}{dz_l dz_k}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\big)\frac{d^2}{dz_l dz_k}\sin\theta_{ijkl} + \frac{d}{dz_k}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dz_k}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.311}$$

$$\frac{d^2}{dx_l^2} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-\big(-\cos\theta_{ijkl} + \cos\theta^0_{ijk}\big)\frac{d^2}{dx_l^2}\cos\theta_{ijkl} - \big(-\sin\theta_{ijkl} + \sin\theta^0_{ijk}\big)\frac{d^2}{dx_l^2}\sin\theta_{ijkl} + \big(\frac{d}{dx_l}\cos\theta_{ijkl}\big)^2 + \big(\frac{d}{dx_l}\sin\theta_{ijkl}\big)^2\big) \tag{.312}$$

$$\frac{d^2}{dy_l dx_l} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_l dx_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_l dx_l}\sin\theta_{ijkl} \\ + \frac{d}{dx_l}\cos\theta_{ijkl}\frac{d}{dy_l}\cos\theta_{ijkl} + \frac{d}{dx_l}\sin\theta_{ijkl}\frac{d}{dy_l}\sin\theta_{ijkl}\big) \tag{.313}$$

$$\frac{d^2}{dz_l dx_l} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l dx_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l dx_l}\sin\theta_{ijkl} \\ + \frac{d}{dx_l}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dx_l}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.314}$$

$$\frac{d^2}{dy_l^2} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dy_l^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dy_l^2}\sin\theta_{ijkl} \\ + (\frac{d}{dy_l}\cos\theta_{ijkl})^2 + (\frac{d}{dy_l}\sin\theta_{ijkl})^2\big) \tag{.315}$$

$$\frac{d^2}{dz_l dy_l} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l dy_l}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l dy_l}\sin\theta_{ijkl} \\ + \frac{d}{dy_l}\cos\theta_{ijkl}\frac{d}{dz_l}\cos\theta_{ijkl} + \frac{d}{dy_l}\sin\theta_{ijkl}\frac{d}{dz_l}\sin\theta_{ijkl}\big) \tag{.316}$$

$$\frac{d^2}{dz_l^2} E_{\text{dihedral},ijkl} = 2k_{\text{dihedral}}\big(-(-\cos\theta_{ijkl} + \cos\theta^0_{ijk})\frac{d^2}{dz_l^2}\cos\theta_{ijkl} - (-\sin\theta_{ijkl} + \sin\theta^0_{ijk})\frac{d^2}{dz_l^2}\sin\theta_{ijkl} \\ + (\frac{d}{dz_l}\cos\theta_{ijkl})^2 + (\frac{d}{dz_l}\sin\theta_{ijkl})^2\big) \tag{.317}$$

The derivatives of $\cos\theta_{ijkl}$ and $\sin\theta_{ijkl}$ with respect to the coordinates x, y and z of the participating atoms i, j, k and l are defined in Section B. The derivatives of $\cos\theta_{ijkl}$ and $\sin\theta_{ijkl}$ are identical, except for interchanging a_ijkl and d_ijkl.

Derivatives of $\cos\theta_{ijkl}$

The first and second order derivatives of $\cos\theta_{ijkl}$ with respect to the coordinates x, y and z of the participating atoms i, j, k and l are:

$$\frac{d}{dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}}{b_{ijk}^2 c_{jkl}} \tag{.318}$$

$$\frac{d}{dy_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}}{b_{ijk}^2 c_{jkl}} \tag{.319}$$

$$\frac{d}{dz_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}}{b_{ijk}^2 c_{jkl}} \tag{.320}$$

$$\frac{d}{dx_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.321}$$

$$\frac{d}{dy_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.322}$$

$$\frac{d}{dz_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.323}$$

$$\frac{d}{dx_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.324}$$

$$\frac{d}{dy_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_k}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.325}$$

$$\frac{d}{dz_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_k}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.326}$$

$$\frac{d}{dx_l}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d}{dx_l}c_{jkl} + c_{jkl}\frac{d}{dx_l}a_{ijkl}}{b_{ijk}c_{jkl}^2} \tag{.327}$$

$$\frac{d}{dy_l}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d}{dy_l}c_{jkl} + c_{jkl}\frac{d}{dy_l}a_{ijkl}}{b_{ijk}c_{jkl}^2} \tag{.328}$$

$$\frac{d}{dz_l}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d}{dz_l}c_{jkl} + c_{jkl}\frac{d}{dz_l}a_{ijkl}}{b_{ijk}c_{jkl}^2} \tag{.329}$$

$$\frac{d^2}{dx_i^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dx_i^2}b_{ijk} + b_{ijk}\frac{d^2}{dx_i^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dx_i}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.330}$$

$$\frac{d^2}{dy_i dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_i dx_i}b_{ijk} + b_{ijk}\frac{d^2}{dy_i dx_i}a_{ijkl} + \frac{d}{dx_i}a_{ijkl}\frac{d}{dy_i}b_{ijk} - \frac{d}{dy_i}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dy_i}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.331}$$

$$\frac{d^2}{dz_i dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_i dx_i}b_{ijk} + b_{ijk}\frac{d^2}{dz_i dx_i}a_{ijkl} + \frac{d}{dx_i}a_{ijkl}\frac{d}{dz_i}b_{ijk} - \frac{d}{dz_i}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dz_i}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.332}$$

$$\frac{d^2}{dx_j dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dx_j dx_i}b_{ijk} + b_{ijk}\frac{d^2}{dx_j dx_i}a_{ijkl} + \frac{d}{dx_i}a_{ijkl}\frac{d}{dx_j}b_{ijk} - \frac{d}{dx_j}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dx_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dx_j}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.333}$$

$$\frac{d^2}{dy_j dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_j dx_i}b_{ijk} + b_{ijk}\frac{d^2}{dy_j dx_i}a_{ijkl} + \frac{d}{dx_i}a_{ijkl}\frac{d}{dy_j}b_{ijk} - \frac{d}{dy_j}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dy_j}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.334}$$

$$\frac{d^2}{dz_j dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_j dx_i}b_{ijk} + b_{ijk}\frac{d^2}{dz_j dx_i}a_{ijkl} + \frac{d}{dx_i}a_{ijkl}\frac{d}{dz_j}b_{ijk} - \frac{d}{dz_j}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dz_j}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.335}$$

$$\frac{d^2}{dx_k dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dx_k dx_i}b_{ijk} + b_{ijk}\frac{d^2}{dx_k dx_i}a_{ijkl} + \frac{d}{dx_i}a_{ijkl}\frac{d}{dx_k}b_{ijk} - \frac{d}{dx_k}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dx_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.336}$$

$$\frac{d^2}{dy_k dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_k dx_i}b_{ijk} + b_{ijk}\frac{d^2}{dy_k dx_i}a_{ijkl} + \frac{d}{dx_i}a_{ijkl}\frac{d}{dy_k}b_{ijk} - \frac{d}{dy_k}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dy_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.337}$$

$$\frac{d^2}{dz_k dx_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_k dx_i}b_{ijk} + b_{ijk}\frac{d^2}{dz_k dx_i}a_{ijkl} + \frac{d}{dx_i}a_{ijkl}\frac{d}{dz_k}b_{ijk} - \frac{d}{dz_k}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.338}$$

$$\frac{d^2}{dx_l dx_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dx_l dx_i}a_{ijkl} - \frac{d}{dx_l}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.339}$$

$$\frac{d^2}{dy_l dx_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dy_l dx_i}a_{ijkl} - \frac{d}{dy_l}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.340}$$

$$\frac{d^2}{dz_l dx_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dz_l dx_i}a_{ijkl} - \frac{d}{dz_l}a_{ijkl}\frac{d}{dx_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dx_i}b_{ijk} + b_{ijk}\frac{d}{dx_i}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.341}$$

$$\frac{d^2}{dy_i^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_i^2}b_{ijk} + b_{ijk}\frac{d^2}{dy_i^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}} - \frac{2\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dy_i}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.342}$$

$$\begin{aligned}\frac{d^2}{dz_i dy_i}\cos\theta_{ijkl} = {} & \frac{-a_{ijkl}\frac{d^2}{dz_i dy_i}b_{ijk} + b_{ijk}\frac{d^2}{dz_i dy_i}a_{ijkl} + \frac{d}{dy_i}a_{ijkl}\frac{d}{dz_i}b_{ijk} - \frac{d}{dz_i}a_{ijkl}\frac{d}{dy_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} \\ & - \frac{2\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dz_i}b_{ijk}}{b_{ijk}^3 c_{jkl}}\end{aligned} \tag{.343}$$

$$\begin{aligned}\frac{d^2}{dx_j dy_i}\cos\theta_{ijkl} = {} & \frac{-a_{ijkl}\frac{d^2}{dy_i dx_j}b_{ijk} + b_{ijk}\frac{d^2}{dy_i dx_j}a_{ijkl} - \frac{d}{dx_j}a_{ijkl}\frac{d}{dy_i}b_{ijk} + \frac{d}{dy_i}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}} \\ & - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dx_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \\ & - \frac{2\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dx_j}b_{ijk}}{b_{ijk}^3 c_{jkl}}\end{aligned} \tag{.344}$$

$$\frac{d^2}{dy_j dy_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_j dy_i}b_{ijk} + b_{ijk}\frac{d^2}{dy_j dy_i}a_{ijkl} + \frac{d}{dy_i}a_{ijkl}\frac{d}{dy_j}b_{ijk} - \frac{d}{dy_j}a_{ijkl}\frac{d}{dy_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dy_j}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.345}$$

$$\frac{d^2}{dz_j dy_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_j dy_i}b_{ijk} + b_{ijk}\frac{d^2}{dz_j dy_i}a_{ijkl} + \frac{d}{dy_i}a_{ijkl}\frac{d}{dz_j}b_{ijk} - \frac{d}{dz_j}a_{ijkl}\frac{d}{dy_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dz_j}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.346}$$

$$\frac{d^2}{dx_k dy_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_i dx_k}b_{ijk} + b_{ijk}\frac{d^2}{dy_i dx_k}a_{ijkl} - \frac{d}{dx_k}a_{ijkl}\frac{d}{dy_i}b_{ijk} + \frac{d}{dy_i}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dx_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.347}$$

$$\frac{d^2}{dy_k dy_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_k dy_i}b_{ijk} + b_{ijk}\frac{d^2}{dy_k dy_i}a_{ijkl} + \frac{d}{dy_i}a_{ijkl}\frac{d}{dy_k}b_{ijk} - \frac{d}{dy_k}a_{ijkl}\frac{d}{dy_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dy_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.348}$$

$$\frac{d^2}{dz_k dy_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_k dy_i}b_{ijk} + b_{ijk}\frac{d^2}{dz_k dy_i}a_{ijkl} + \frac{d}{dy_i}a_{ijkl}\frac{d}{dz_k}b_{ijk} - \frac{d}{dz_k}a_{ijkl}\frac{d}{dy_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.349}$$

$$\frac{d^2}{dx_l dy_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dy_i dx_l}a_{ijkl} - \frac{d}{dx_l}a_{ijkl}\frac{d}{dy_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.350}$$

$$\frac{d^2}{dy_l dy_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dy_l dy_i}a_{ijkl} - \frac{d}{dy_l}a_{ijkl}\frac{d}{dy_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.351}$$

$$\frac{d^2}{dz_l dy_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dz_l dy_i}a_{ijkl} - \frac{d}{dz_l}a_{ijkl}\frac{d}{dy_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dy_i}b_{ijk} + b_{ijk}\frac{d}{dy_i}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.352}$$

$$\frac{d^2}{dz_i^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_i^2}b_{ijk} + b_{ijk}\frac{d^2}{dz_i^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}} - \frac{2\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dz_i}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.353}$$

$$\begin{aligned}\frac{d^2}{dx_j dz_i}\cos\theta_{ijkl} = &\frac{-a_{ijkl}\frac{d^2}{dz_i dx_j}b_{ijk} + b_{ijk}\frac{d^2}{dz_i dx_j}a_{ijkl} - \frac{d}{dx_j}a_{ijkl}\frac{d}{dz_i}b_{ijk} + \frac{d}{dz_i}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}} \\ &- \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dx_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \\ &- \frac{2\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dx_j}b_{ijk}}{b_{ijk}^3 c_{jkl}}\end{aligned} \tag{.354}$$

$$\begin{aligned}\frac{d^2}{dy_j dz_i}\cos\theta_{ijkl} = &\frac{-a_{ijkl}\frac{d^2}{dz_i dy_j}b_{ijk} + b_{ijk}\frac{d^2}{dz_i dy_j}a_{ijkl} - \frac{d}{dy_j}a_{ijkl}\frac{d}{dz_i}b_{ijk} + \frac{d}{dz_i}a_{ijkl}\frac{d}{dy_j}b_{ijk}}{b_{ijk}^2 c_{jkl}} \\ &- \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \\ &- \frac{2\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dy_j}b_{ijk}}{b_{ijk}^3 c_{jkl}}\end{aligned} \tag{.355}$$

$$\begin{aligned}\frac{d^2}{dz_j dz_i}\cos\theta_{ijkl} = &\frac{-a_{ijkl}\frac{d^2}{dz_j dz_i}b_{ijk} + b_{ijk}\frac{d^2}{dz_j dz_i}a_{ijkl} + \frac{d}{dz_i}a_{ijkl}\frac{d}{dz_j}b_{ijk} - \frac{d}{dz_j}a_{ijkl}\frac{d}{dz_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} \\ &- \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \\ &- \frac{2\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dz_j}b_{ijk}}{b_{ijk}^3 c_{jkl}}\end{aligned} \tag{.356}$$

$$\frac{d^2}{dx_k dz_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_i dx_k}b_{ijk} + b_{ijk}\frac{d^2}{dz_i dx_k}a_{ijkl} - \frac{d}{dx_k}a_{ijkl}\frac{d}{dz_i}b_{ijk} + \frac{d}{dz_i}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dx_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.357}$$

$$\frac{d^2}{dy_k dz_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_i dy_k}b_{ijk} + b_{ijk}\frac{d^2}{dz_i dy_k}a_{ijkl} - \frac{d}{dy_k}a_{ijkl}\frac{d}{dz_i}b_{ijk} + \frac{d}{dz_i}a_{ijkl}\frac{d}{dy_k}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dy_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.358}$$

$$\frac{d^2}{dz_k dz_i}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_k dz_i}b_{ijk} + b_{ijk}\frac{d^2}{dz_k dz_i}a_{ijkl} + \frac{d}{dz_i}a_{ijkl}\frac{d}{dz_k}b_{ijk} - \frac{d}{dz_k}a_{ijkl}\frac{d}{dz_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}} \tag{.359}$$

$$\frac{d^2}{dx_l dz_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dz_i dx_l}a_{ijkl} - \frac{d}{dx_l}a_{ijkl}\frac{d}{dz_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.360}$$

$$\frac{d^2}{dy_l dz_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dz_i dy_l}a_{ijkl} - \frac{d}{dy_l}a_{ijkl}\frac{d}{dz_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.361}$$

$$\frac{d^2}{dz_l dz_i}\cos\theta_{ijkl} = \frac{b_{ijk}\frac{d^2}{dz_l dz_i}a_{ijkl} - \frac{d}{dz_l}a_{ijkl}\frac{d}{dz_i}b_{ijk}}{b_{ijk}^2 c_{jkl}} - \frac{\left(-a_{ijkl}\frac{d}{dz_i}b_{ijk} + b_{ijk}\frac{d}{dz_i}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.362}$$

$$\begin{aligned}\frac{d^2}{dx_j^2}\cos\theta_{ijkl} = {} & \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dx_j^2}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dx_j^2}b_{ijk} - 2a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dx_j}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dx_j^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \\ & - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dx_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} \\ & - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dx_j}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}\end{aligned} \tag{.363}$$

$$\begin{aligned}\frac{d^2}{dy_j dx_j}\cos\theta_{ijkl} = {} & \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_j dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dy_j dx_j}b_{ijk} - a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dx_j}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_j dx_j}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \\ & + \frac{b_{ijk}\frac{d}{dx_j}a_{ijkl}\frac{d}{dy_j}c_{jkl} - b_{ijk}\frac{d}{dy_j}a_{ijkl}\frac{d}{dx_j}c_{jkl} + c_{jkl}\frac{d}{dx_j}a_{ijkl}\frac{d}{dy_j}b_{ijk} - c_{jkl}\frac{d}{dy_j}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \\ & - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dy_j}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}\end{aligned} \tag{.364}$$

$$\frac{d^2}{dz_j dx_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_j dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_j dx_j}b_{ijk} - a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dx_j}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_j dx_j}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.365}$$
$$+ \frac{b_{ijk}\frac{d}{dx_j}a_{ijkl}\frac{d}{dz_j}c_{jkl} - b_{ijk}\frac{d}{dz_j}a_{ijkl}\frac{d}{dx_j}c_{jkl} + c_{jkl}\frac{d}{dx_j}a_{ijkl}\frac{d}{dz_j}b_{ijk} - c_{jkl}\frac{d}{dz_j}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dz_j}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dx_k dx_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dx_k dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dx_k dx_j}b_{ijk} - a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dx_j}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dx_k dx_j}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.366}$$
$$+ \frac{b_{ijk}\frac{d}{dx_j}a_{ijkl}\frac{d}{dx_k}c_{jkl} - b_{ijk}\frac{d}{dx_k}a_{ijkl}\frac{d}{dx_j}c_{jkl} + c_{jkl}\frac{d}{dx_j}a_{ijkl}\frac{d}{dx_k}b_{ijk} - c_{jkl}\frac{d}{dx_k}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dx_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dy_k dx_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_k dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dy_k dx_j}b_{ijk} - a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dx_j}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_k dx_j}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.367}$$
$$+ \frac{b_{ijk}\frac{d}{dx_j}a_{ijkl}\frac{d}{dy_k}c_{jkl} - b_{ijk}\frac{d}{dy_k}a_{ijkl}\frac{d}{dx_j}c_{jkl} + c_{jkl}\frac{d}{dx_j}a_{ijkl}\frac{d}{dy_k}b_{ijk} - c_{jkl}\frac{d}{dy_k}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dy_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dz_k dx_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_k dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_k dx_j}b_{ijk} - a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dx_j}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_k dx_j}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.368}$$

$$+ \frac{b_{ijk}\frac{d}{dx_j}a_{ijkl}\frac{d}{dz_k}c_{jkl} - b_{ijk}\frac{d}{dz_k}a_{ijkl}\frac{d}{dx_j}c_{jkl} + c_{jkl}\frac{d}{dx_j}a_{ijkl}\frac{d}{dz_k}b_{ijk} - c_{jkl}\frac{d}{dz_k}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dx_l dx_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dx_l dx_j}c_{jkl} - a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dx_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dx_l dx_j}a_{ijkl} + b_{ijk}\frac{d}{dx_j}a_{ijkl}\frac{d}{dx_l}c_{jkl} - b_{ijk}\frac{d}{dx_l}a_{ijkl}\frac{d}{dx_j}c_{jkl} - c_{jkl}\frac{d}{dx_l}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.369}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dy_l dx_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_l dx_j}c_{jkl} - a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dy_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_l dx_j}a_{ijkl} + b_{ijk}\frac{d}{dx_j}a_{ijkl}\frac{d}{dy_l}c_{jkl} - b_{ijk}\frac{d}{dy_l}a_{ijkl}\frac{d}{dx_j}c_{jkl} - c_{jkl}\frac{d}{dy_l}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.370}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dz_l dx_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_l dx_j}c_{jkl} - a_{ijkl}\frac{d}{dx_j}b_{ijk}\frac{d}{dz_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_l dx_j}a_{ijkl} + b_{ijk}\frac{d}{dx_j}a_{ijkl}\frac{d}{dz_l}c_{jkl} - b_{ijk}\frac{d}{dz_l}a_{ijkl}\frac{d}{dx_j}c_{jkl} - c_{jkl}\frac{d}{dz_l}a_{ijkl}\frac{d}{dx_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.371}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_j}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dy_j^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_j^2}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dy_j^2}b_{ijk} - 2a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dy_j}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_j^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.372}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dy_j}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dz_j dy_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_j dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_j dy_j}b_{ijk} - a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.373}$$

$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dz_j dy_j}a_{ijkl} + b_{ijk}\frac{d}{dy_j}a_{ijkl}\frac{d}{dz_j}c_{jkl} - b_{ijk}\frac{d}{dz_j}a_{ijkl}\frac{d}{dy_j}c_{jkl} + c_{jkl}\frac{d}{dy_j}a_{ijkl}\frac{d}{dz_j}b_{ijk} - c_{jkl}\frac{d}{dz_j}a_{ijkl}\frac{d}{dy_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dz_j}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dx_k dy_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_j dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dy_j dx_k}b_{ijk} - a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.374}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dy_j dx_k}a_{ijkl} - b_{ijk}\frac{d}{dx_k}a_{ijkl}\frac{d}{dy_j}c_{jkl} + b_{ijk}\frac{d}{dy_j}a_{ijkl}\frac{d}{dx_k}c_{jkl} - c_{jkl}\frac{d}{dx_k}a_{ijkl}\frac{d}{dy_j}b_{ijk} + c_{jkl}\frac{d}{dy_j}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dx_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dy_k dy_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_k dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dy_k dy_j}b_{ijk} - a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.375}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dy_k dy_j}a_{ijkl} + b_{ijk}\frac{d}{dy_j}a_{ijkl}\frac{d}{dy_k}c_{jkl} - b_{ijk}\frac{d}{dy_k}a_{ijkl}\frac{d}{dy_j}c_{jkl} + c_{jkl}\frac{d}{dy_j}a_{ijkl}\frac{d}{dy_k}b_{ijk} - c_{jkl}\frac{d}{dy_k}a_{ijkl}\frac{d}{dy_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dy_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dz_k dy_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_k dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_k dy_j}b_{ijk} - a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.376}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dz_k dy_j}a_{ijkl} + b_{ijk}\frac{d}{dy_j}a_{ijkl}\frac{d}{dz_k}c_{jkl} - b_{ijk}\frac{d}{dz_k}a_{ijkl}\frac{d}{dy_j}c_{jkl} + c_{jkl}\frac{d}{dy_j}a_{ijkl}\frac{d}{dz_k}b_{ijk} - c_{jkl}\frac{d}{dz_k}a_{ijkl}\frac{d}{dy_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dx_l dy_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_j dx_l}c_{jkl} - a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dx_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_j dx_l}a_{ijkl} - b_{ijk}\frac{d}{dx_l}a_{ijkl}\frac{d}{dy_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} + \frac{b_{ijk}\frac{d}{dy_j}a_{ijkl}\frac{d}{dx_l}c_{jkl} - c_{jkl}\frac{d}{dx_l}a_{ijkl}\frac{d}{dy_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} \tag{.377}$$

$$\frac{d^2}{dy_l dy_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_l dy_j}c_{jkl} - a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dy_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_l dy_j}a_{ijkl} + b_{ijk}\frac{d}{dy_j}a_{ijkl}\frac{d}{dy_l}c_{jkl} - b_{ijk}\frac{d}{dy_l}a_{ijkl}\frac{d}{dy_j}c_{jkl} - c_{jkl}\frac{d}{dy_l}a_{ijkl}\frac{d}{dy_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} \tag{.378}$$

$$\frac{d^2}{dz_l dy_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_l dy_j}c_{jkl} - a_{ijkl}\frac{d}{dy_j}b_{ijk}\frac{d}{dz_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_l dy_j}a_{ijkl} + b_{ijk}\frac{d}{dy_j}a_{ijkl}\frac{d}{dz_l}c_{jkl} - b_{ijk}\frac{d}{dz_l}a_{ijkl}\frac{d}{dy_j}c_{jkl} - c_{jkl}\frac{d}{dz_l}a_{ijkl}\frac{d}{dy_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_j}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} \tag{.379}$$

$$\frac{d^2}{dz_j^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_j^2}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_j^2}b_{ijk} - 2a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dz_j}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_j^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dz_j}b_{ijk}}{b_{ijk}^3 c_{jkl}^2} \tag{.380}$$

$$\frac{d^2}{dx_k dz_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_j dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_j dx_k}b_{ijk} - a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.381}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dz_j dx_k}a_{ijkl} - b_{ijk}\frac{d}{dx_k}a_{ijkl}\frac{d}{dz_j}c_{jkl} + b_{ijk}\frac{d}{dz_j}a_{ijkl}\frac{d}{dx_k}c_{jkl} - c_{jkl}\frac{d}{dx_k}a_{ijkl}\frac{d}{dz_j}b_{ijk} + c_{jkl}\frac{d}{dz_j}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dx_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dy_k dz_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_j dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_j dy_k}b_{ijk} - a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.382}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dz_j dy_k}a_{ijkl} - b_{ijk}\frac{d}{dy_k}a_{ijkl}\frac{d}{dz_j}c_{jkl} + b_{ijk}\frac{d}{dz_j}a_{ijkl}\frac{d}{dy_k}c_{jkl} - c_{jkl}\frac{d}{dy_k}a_{ijkl}\frac{d}{dz_j}b_{ijk} + c_{jkl}\frac{d}{dz_j}a_{ijkl}\frac{d}{dy_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dy_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dz_k dz_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_k dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_k dz_j}b_{ijk} - a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.383}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dz_k dz_j}a_{ijkl} + b_{ijk}\frac{d}{dz_j}a_{ijkl}\frac{d}{dz_k}c_{jkl} - b_{ijk}\frac{d}{dz_k}a_{ijkl}\frac{d}{dz_j}c_{jkl} + c_{jkl}\frac{d}{dz_j}a_{ijkl}\frac{d}{dz_k}b_{ijk} - c_{jkl}\frac{d}{dz_k}a_{ijkl}\frac{d}{dz_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dx_l dz_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_j dx_l}c_{jkl} - a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dx_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_j dx_l}a_{ijkl} - b_{ijk}\frac{d}{dx_l}a_{ijkl}\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} + \frac{b_{ijk}\frac{d}{dz_j}a_{ijkl}\frac{d}{dx_l}c_{jkl} - c_{jkl}\frac{d}{dx_l}a_{ijkl}\frac{d}{dz_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} \tag{.384}$$

$$\frac{d^2}{dy_l dz_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_j dy_l}c_{jkl} - a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dy_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_j dy_l}a_{ijkl} - b_{ijk}\frac{d}{dy_l}a_{ijkl}\frac{d}{dz_j}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} + \frac{b_{ijk}\frac{d}{dz_j}a_{ijkl}\frac{d}{dy_l}c_{jkl} - c_{jkl}\frac{d}{dy_l}a_{ijkl}\frac{d}{dz_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} \tag{.385}$$

$$\frac{d^2}{dz_l dz_j}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_l dz_j}c_{jkl} - a_{ijkl}\frac{d}{dz_j}b_{ijk}\frac{d}{dz_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_l dz_j}a_{ijkl} + b_{ijk}\frac{d}{dz_j}a_{ijkl}\frac{d}{dz_l}c_{jkl} - b_{ijk}\frac{d}{dz_l}a_{ijkl}\frac{d}{dz_j}c_{jkl} - c_{jkl}\frac{d}{dz_l}a_{ijkl}\frac{d}{dz_j}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_j}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_j}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_j}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} \tag{.386}$$

$$\frac{d^2}{dx_k^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dx_k^2}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dx_k^2}b_{ijk} - 2a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dx_k}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dx_k^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dx_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2} \tag{.387}$$

$$\frac{d^2}{dy_k dx_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_k dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dy_k dx_k}b_{ijk} - a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.388}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dy_k dx_k}a_{ijkl} + b_{ijk}\frac{d}{dx_k}a_{ijkl}\frac{d}{dy_k}c_{jkl} - b_{ijk}\frac{d}{dy_k}a_{ijkl}\frac{d}{dx_k}c_{jkl} + c_{jkl}\frac{d}{dx_k}a_{ijkl}\frac{d}{dy_k}b_{ijk} - c_{jkl}\frac{d}{dy_k}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dy_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dz_k dx_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_k dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_k dx_k}b_{ijk} - a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dx_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.389}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dz_k dx_k}a_{ijkl} + b_{ijk}\frac{d}{dx_k}a_{ijkl}\frac{d}{dz_k}c_{jkl} - b_{ijk}\frac{d}{dz_k}a_{ijkl}\frac{d}{dx_k}c_{jkl} + c_{jkl}\frac{d}{dx_k}a_{ijkl}\frac{d}{dz_k}b_{ijk} - c_{jkl}\frac{d}{dz_k}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dx_l dx_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dx_l dx_k}c_{jkl} - a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dx_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dx_l dx_k}a_{ijkl} + b_{ijk}\frac{d}{dx_k}a_{ijkl}\frac{d}{dx_l}c_{jkl} - b_{ijk}\frac{d}{dx_l}a_{ijkl}\frac{d}{dx_k}c_{jkl} - c_{jkl}\frac{d}{dx_l}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.390}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dy_l dx_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_l dx_k}c_{jkl} - a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dy_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_l dx_k}a_{ijkl} + b_{ijk}\frac{d}{dx_k}a_{ijkl}\frac{d}{dy_l}c_{jkl} - b_{ijk}\frac{d}{dy_l}a_{ijkl}\frac{d}{dx_k}c_{jkl} - c_{jkl}\frac{d}{dy_l}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.391}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dz_l dx_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_l dx_k}c_{jkl} - a_{ijkl}\frac{d}{dx_k}b_{ijk}\frac{d}{dz_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_l dx_k}a_{ijkl} + b_{ijk}\frac{d}{dx_k}a_{ijkl}\frac{d}{dz_l}c_{jkl} - b_{ijk}\frac{d}{dz_l}a_{ijkl}\frac{d}{dx_k}c_{jkl} - c_{jkl}\frac{d}{dz_l}a_{ijkl}\frac{d}{dx_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.392}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dx_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dx_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dx_k}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dy_k^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_k^2}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dy_k^2}b_{ijk} - 2a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dy_k}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_k^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.393}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_k}a_{ijkl}\right)\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_k}a_{ijkl}\right)\frac{d}{dy_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dz_k dy_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_k dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_k dy_k}b_{ijk} - a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dy_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.394}$$
$$+ \frac{b_{ijk}c_{jkl}\frac{d^2}{dz_k dy_k}a_{ijkl} + b_{ijk}\frac{d}{dy_k}a_{ijkl}\frac{d}{dz_k}c_{jkl} - b_{ijk}\frac{d}{dz_k}a_{ijkl}\frac{d}{dy_k}c_{jkl} + c_{jkl}\frac{d}{dy_k}a_{ijkl}\frac{d}{dz_k}b_{ijk} - c_{jkl}\frac{d}{dz_k}a_{ijkl}\frac{d}{dy_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_k}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_k}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dx_l dy_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_k dx_l}c_{jkl} - a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dx_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_k dx_l}a_{ijkl} - b_{ijk}\frac{d}{dx_l}a_{ijkl}\frac{d}{dy_k}c_{jkl} + b_{ijk}\frac{d}{dy_k}a_{ijkl}\frac{d}{dx_l}c_{jkl} - c_{jkl}\frac{d}{dx_l}a_{ijkl}\frac{d}{dy_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.395}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_k}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dy_l dy_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dy_l dy_k}c_{jkl} - a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dy_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dy_l dy_k}a_{ijkl} + b_{ijk}\frac{d}{dy_k}a_{ijkl}\frac{d}{dy_l}c_{jkl} - b_{ijk}\frac{d}{dy_l}a_{ijkl}\frac{d}{dy_k}c_{jkl} - c_{jkl}\frac{d}{dy_l}a_{ijkl}\frac{d}{dy_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.396}$$
$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_k}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dz_l dy_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_l dy_k}c_{jkl} - a_{ijkl}\frac{d}{dy_k}b_{ijk}\frac{d}{dz_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_l dy_k}a_{ijkl} + b_{ijk}\frac{d}{dy_k}a_{ijkl}\frac{d}{dz_l}c_{jkl} - b_{ijk}\frac{d}{dz_l}a_{ijkl}\frac{d}{dy_k}c_{jkl} - c_{jkl}\frac{d}{dz_l}a_{ijkl}\frac{d}{dy_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.397}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dy_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dy_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dy_k}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dz_k^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_k^2}c_{jkl} - a_{ijkl}c_{jkl}\frac{d^2}{dz_k^2}b_{ijk} - 2a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dz_k}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_k^2}a_{ijkl}}{b_{ijk}^2 c_{jkl}^2} \tag{.398}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_k}a_{ijkl}\right)\frac{d}{dz_k}c_{jkl}}{b_{ijk}^2 c_{jkl}^3} - \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_k}a_{ijkl}\right)\frac{d}{dz_k}b_{ijk}}{b_{ijk}^3 c_{jkl}^2}$$

$$\frac{d^2}{dx_l dz_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_k dx_l}c_{jkl} - a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dx_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_k dx_l}a_{ijkl} - b_{ijk}\frac{d}{dx_l}a_{ijkl}\frac{d}{dz_k}c_{jkl} + b_{ijk}\frac{d}{dz_k}a_{ijkl}\frac{d}{dx_l}c_{jkl} - c_{jkl}\frac{d}{dx_l}a_{ijkl}\frac{d}{dz_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.399}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_k}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dy_l dz_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_k dy_l}c_{jkl} - a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dy_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_k dy_l}a_{ijkl} - b_{ijk}\frac{d}{dy_l}a_{ijkl}\frac{d}{dz_k}c_{jkl} + b_{ijk}\frac{d}{dz_k}a_{ijkl}\frac{d}{dy_l}c_{jkl} - c_{jkl}\frac{d}{dy_l}a_{ijkl}\frac{d}{dz_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.400}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_k}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dz_l dz_k}\cos\theta_{ijkl} = \frac{-a_{ijkl}b_{ijk}\frac{d^2}{dz_l dz_k}c_{jkl} - a_{ijkl}\frac{d}{dz_k}b_{ijk}\frac{d}{dz_l}c_{jkl} + b_{ijk}c_{jkl}\frac{d^2}{dz_l dz_k}a_{ijkl} + b_{ijk}\frac{d}{dz_k}a_{ijkl}\frac{d}{dz_l}c_{jkl} - b_{ijk}\frac{d}{dz_l}a_{ijkl}\frac{d}{dz_k}c_{jkl} - c_{jkl}\frac{d}{dz_l}a_{ijkl}\frac{d}{dz_k}b_{ijk}}{b_{ijk}^2 c_{jkl}^2} \tag{.401}$$

$$- \frac{2\left(-a_{ijkl}b_{ijk}\frac{d}{dz_k}c_{jkl} - a_{ijkl}c_{jkl}\frac{d}{dz_k}b_{ijk} + b_{ijk}c_{jkl}\frac{d}{dz_k}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}^2 c_{jkl}^3}$$

$$\frac{d^2}{dx_l^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dx_l^2}c_{jkl} + c_{jkl}\frac{d^2}{dx_l^2}a_{ijkl}}{b_{ijk}c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_l}c_{jkl} + c_{jkl}\frac{d}{dx_l}a_{ijkl}\right)\frac{d}{dx_l}c_{jkl}}{b_{ijk}c_{jkl}^3} \tag{.402}$$

$$\frac{d^2}{dy_l dx_l}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_l dx_l}c_{jkl} + c_{jkl}\frac{d^2}{dy_l dx_l}a_{ijkl} + \frac{d}{dx_l}a_{ijkl}\frac{d}{dy_l}c_{jkl} - \frac{d}{dy_l}a_{ijkl}\frac{d}{dx_l}c_{jkl}}{b_{ijk}c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_l}c_{jkl} + c_{jkl}\frac{d}{dx_l}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}c_{jkl}^3} \tag{.403}$$

$$\frac{d^2}{dz_l dx_l}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_l dx_l}c_{jkl} + c_{jkl}\frac{d^2}{dz_l dx_l}a_{ijkl} + \frac{d}{dx_l}a_{ijkl}\frac{d}{dz_l}c_{jkl} - \frac{d}{dz_l}a_{ijkl}\frac{d}{dx_l}c_{jkl}}{b_{ijk}c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dx_l}c_{jkl} + c_{jkl}\frac{d}{dx_l}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}c_{jkl}^3} \tag{.404}$$

$$\frac{d^2}{dy_l^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dy_l^2}c_{jkl} + c_{jkl}\frac{d^2}{dy_l^2}a_{ijkl}}{b_{ijk}c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dy_l}c_{jkl} + c_{jkl}\frac{d}{dy_l}a_{ijkl}\right)\frac{d}{dy_l}c_{jkl}}{b_{ijk}c_{jkl}^3} \tag{.405}$$

$$\frac{d^2}{dz_l dy_l}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_l dy_l}c_{jkl} + c_{jkl}\frac{d^2}{dz_l dy_l}a_{ijkl} + \frac{d}{dy_l}a_{ijkl}\frac{d}{dz_l}c_{jkl} - \frac{d}{dz_l}a_{ijkl}\frac{d}{dy_l}c_{jkl}}{b_{ijk}c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dy_l}c_{jkl} + c_{jkl}\frac{d}{dy_l}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}c_{jkl}^3} \tag{.406}$$

$$\frac{d^2}{dz_l^2}\cos\theta_{ijkl} = \frac{-a_{ijkl}\frac{d^2}{dz_l^2}c_{jkl} + c_{jkl}\frac{d^2}{dz_l^2}a_{ijkl}}{b_{ijk}c_{jkl}^2} - \frac{2\left(-a_{ijkl}\frac{d}{dz_l}c_{jkl} + c_{jkl}\frac{d}{dz_l}a_{ijkl}\right)\frac{d}{dz_l}c_{jkl}}{b_{ijk}c_{jkl}^3} \tag{.407}$$

The derivatives of a_{ijkl} with respect to the coordinates x, y and z of the participating atoms i, j, k and l are defined in Section B. The derivatives of $\cos\theta_{ijkl}$ are equivalent by replacing a_{ijkl} with d_{ijkl} (see Section B).

Derivatives of a_{ijkl}

The first and second order derivatives of a_{ijkl} with respect to the coordinates x, y and z of the participating atoms i,j,k and l are:

$$\frac{d}{dx_i}a_{ijkl} = -y_{kj}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right) - z_{kj}\left(x_{kj}z_{lk} - x_{lk}z_{kj}\right) \tag{.408}$$

$$\frac{d}{dy_i}a_{ijkl} = x_{kj}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right) - z_{kj}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) \tag{.409}$$

$$\frac{d}{dz_i}a_{ijkl} = x_{kj}\left(x_{kj}z_{lk} - x_{lk}z_{kj}\right) + y_{kj}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) \tag{.410}$$

$$\frac{d}{dx_j}a_{ijkl} = -y_{lk}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - z_{lk}\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right) - \left(y_i - y_k\right)\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right) - \left(z_i - z_k\right)\left(x_{kj}z_{lk} - x_{lk}z_{kj}\right) \tag{.411}$$

$$\frac{d}{dy_j}a_{ijkl} = x_{lk}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - z_{lk}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) + \left(x_i - x_k\right)\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right) - \left(z_i - z_k\right)\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) \tag{.412}$$

$$\frac{d}{dz_j}a_{ijkl} = x_{lk}\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right) + y_{lk}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) + \left(x_i - x_k\right)\left(x_{kj}z_{lk} - x_{lk}z_{kj}\right) + \left(y_i - y_k\right)\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) \tag{.413}$$

$$\frac{d}{dx_k}a_{ijkl} = -y_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right) - z_{ji}\left(x_{kj}z_{lk} - x_{lk}z_{kj}\right) - \left(y_j - y_l\right)\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - \left(z_j - z_l\right)\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right) \tag{.414}$$

$$\frac{d}{dy_k}a_{ijkl} = x_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right) - z_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + \left(x_j - x_l\right)\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - \left(z_j - z_l\right)\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) \tag{.415}$$

$$\frac{d}{dz_k}a_{ijkl} = x_{ji}\left(x_{kj}z_{lk} - x_{lk}z_{kj}\right) + y_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + \left(x_j - x_l\right)\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right) + \left(y_j - y_l\right)\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) \tag{.416}$$

$$\frac{d}{dx_l}a_{ijkl} = -y_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - z_{kj}\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right) \tag{.417}$$

$$\frac{d}{dy_l}a_{ijkl} = x_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - z_{kj}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) \tag{.418}$$

$$\frac{d}{dz_l}a_{ijkl} = x_{kj}\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right) + y_{kj}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) \tag{.419}$$

$$\frac{d^2}{dx_i^2}a_{ijkl} = 0 \tag{.420}$$

$$\frac{d^2}{dy_i dx_i}a_{ijkl} = 0 \tag{.421}$$

$$\frac{d^2}{dz_i dx_i}a_{ijkl} = 0 \tag{.422}$$

$$\frac{d^2}{dx_j dx_i}a_{ijkl} = y_{kj}y_{lk} + z_{kj}z_{lk} \tag{.423}$$

$$\frac{d^2}{dy_j dx_i}a_{ijkl} = x_{kj}y_{lk} - 2x_{lk}y_{kj} \tag{.424}$$

$$\frac{d^2}{dz_j dx_i}a_{ijkl} = x_{kj}z_{lk} - 2x_{lk}z_{kj} \tag{.425}$$

$$\frac{d^2}{dx_k dx_i}a_{ijkl} = -y_{kj}\left(-y_j + y_l\right) - z_{kj}\left(-z_j + z_l\right) \tag{.426}$$

$$\frac{d^2}{dy_k dx_i}a_{ijkl} = -x_{kj}y_{lk} + x_{lk}y_{kj} - y_{kj}\left(x_j - x_l\right) \tag{.427}$$

$$\frac{d^2}{dz_k dx_i}a_{ijkl} = -x_{kj}z_{lk} + x_{lk}z_{kj} - z_{kj}\left(x_j - x_l\right) \tag{.428}$$

$$\frac{d^2}{dx_l dx_i} a_{ijkl} = y_{kj}^2 + z_{kj}^2 \tag{.429}$$

$$\frac{d^2}{dy_l dx_i} a_{ijkl} = -x_{kj} y_{kj} \tag{.430}$$

$$\frac{d^2}{dz_l dx_i} a_{ijkl} = -x_{kj} z_{kj} \tag{.431}$$

$$\frac{d^2}{dy_i^2} a_{ijkl} = 0 \tag{.432}$$

$$\frac{d^2}{dz_i dy_i} a_{ijkl} = 0 \tag{.433}$$

$$\frac{d^2}{dx_j dy_i} a_{ijkl} = -2x_{kj} y_{lk} + x_{lk} y_{kj} \tag{.434}$$

$$\frac{d^2}{dy_j dy_i} a_{ijkl} = x_{kj} x_{lk} + z_{kj} z_{lk} \tag{.435}$$

$$\frac{d^2}{dz_j dy_i} a_{ijkl} = y_{kj} z_{lk} - 2y_{lk} z_{kj} \tag{.436}$$

$$\frac{d^2}{dx_k dy_i} a_{ijkl} = x_{kj} y_{lk} + x_{kj} \left(-y_j + y_l\right) - x_{lk} y_{kj} \tag{.437}$$

$$\frac{d^2}{dy_k dy_i} a_{ijkl} = x_{kj} \left(x_j - x_l\right) - z_{kj} \left(-z_j + z_l\right) \tag{.438}$$

$$\frac{d^2}{dz_k dy_i} a_{ijkl} = -y_{kj} z_{lk} + y_{lk} z_{kj} - z_{kj} \left(y_j - y_l\right) \tag{.439}$$

$$\frac{d^2}{dx_l dy_i} a_{ijkl} = -x_{kj} y_{kj} \tag{.440}$$

$$\frac{d^2}{dy_l dy_i} a_{ijkl} = x_{kj}^2 + z_{kj}^2 \tag{.441}$$

$$\frac{d^2}{dz_l dy_i} a_{ijkl} = -y_{kj} z_{kj} \tag{.442}$$

$$\frac{d^2}{dz_i^2} a_{ijkl} = 0 \tag{.443}$$

$$\frac{d^2}{dx_j dz_i} a_{ijkl} = -2x_{kj} z_{lk} + x_{lk} z_{kj} \tag{.444}$$

$$\frac{d^2}{dy_j dz_i} a_{ijkl} = -2y_{kj} z_{lk} + y_{lk} z_{kj} \tag{.445}$$

$$\frac{d^2}{dz_j dz_i} a_{ijkl} = x_{kj} x_{lk} + y_{kj} y_{lk} \tag{.446}$$

$$\frac{d^2}{dx_k dz_i} a_{ijkl} = x_{kj} z_{lk} + x_{kj} \left(-z_j + z_l\right) - x_{lk} z_{kj} \tag{.447}$$

$$\frac{d^2}{dy_k dz_i} a_{ijkl} = y_{kj} z_{lk} + y_{kj} \left(-z_j + z_l\right) - y_{lk} z_{kj} \tag{.448}$$

$$\frac{d^2}{dz_k dz_i} a_{ijkl} = x_{kj} \left(x_j - x_l\right) + y_{kj} \left(y_j - y_l\right) \tag{.449}$$

$$\frac{d^2}{dx_l dz_i} a_{ijkl} = -x_{kj} z_{kj} \tag{.450}$$

$$\frac{d^2}{dy_l dz_i} a_{ijkl} = -y_{kj} z_{kj} \tag{.451}$$

$$\frac{d^2}{dz_l dz_i} a_{ijkl} = x_{kj}^2 + y_{kj}^2 \tag{.452}$$

$$\frac{d^2}{dx_j^2} a_{ijkl} = -y_{lk} \left(-y_i + y_k\right) + y_{lk} \left(y_i - y_k\right) - z_{lk} \left(-z_i + z_k\right) + z_{lk} \left(z_i - z_k\right) \tag{.453}$$

$$\frac{d^2}{dy_j dx_j} a_{ijkl} = -x_{lk} \left(y_i - y_k\right) - y_{lk} \left(x_i - x_k\right) \tag{.454}$$

$$\frac{d^2}{dz_j dx_j} a_{ijkl} = -x_{lk}\left(z_i - z_k\right) - z_{lk}\left(x_i - x_k\right) \tag{.455}$$

$$\frac{d^2}{dx_k dx_j} a_{ijkl} = y_{ji}y_{lk} + z_{ji}z_{lk} - \left(y_i - y_k\right)\left(-y_j + y_l\right) - \left(z_i - z_k\right)\left(-z_j + z_l\right) \tag{.456}$$

$$\frac{d^2}{dy_k dx_j} a_{ijkl} = x_{ji}y_{kj} - x_{ji}y_{lk} - x_{kj}y_{ji} + x_{kj}y_{lk} - x_{lk}y_{kj} - \left(x_j - x_l\right)\left(y_i - y_k\right) \tag{.457}$$

$$\frac{d^2}{dz_k dx_j} a_{ijkl} = x_{ji}z_{kj} - x_{ji}z_{lk} - x_{kj}z_{ji} + x_{kj}z_{lk} - x_{lk}z_{kj} - \left(x_j - x_l\right)\left(z_i - z_k\right) \tag{.458}$$

$$\frac{d^2}{dx_l dx_j} a_{ijkl} = y_{kj}\left(y_i - y_k\right) + z_{kj}\left(z_i - z_k\right) \tag{.459}$$

$$\frac{d^2}{dy_l dx_j} a_{ijkl} = -x_{ji}y_{kj} + x_{kj}y_{ji} - x_{kj}\left(y_i - y_k\right) \tag{.460}$$

$$\frac{d^2}{dz_l dx_j} a_{ijkl} = -x_{ji}z_{kj} + x_{kj}z_{ji} - x_{kj}\left(z_i - z_k\right) \tag{.461}$$

$$\frac{d^2}{dy_j^2} a_{ijkl} = 2x_{lk}\left(x_i - x_k\right) - z_{lk}\left(-z_i + z_k\right) + z_{lk}\left(z_i - z_k\right) \tag{.462}$$

$$\frac{d^2}{dz_j dy_j} a_{ijkl} = -y_{lk}\left(z_i - z_k\right) - z_{lk}\left(y_i - y_k\right) \tag{.463}$$

$$\frac{d^2}{dx_k dy_j} a_{ijkl} = -x_{ji}y_{kj} + x_{kj}y_{ji} - x_{kj}y_{lk} - x_{lk}y_{ji} + x_{lk}y_{kj} + \left(x_i - x_k\right)\left(-y_j + y_l\right) \tag{.464}$$

$$\frac{d^2}{dy_k dy_j} a_{ijkl} = x_{ji}x_{lk} + z_{ji}z_{lk} + \left(x_i - x_k\right)\left(x_j - x_l\right) - \left(z_i - z_k\right)\left(-z_j + z_l\right) \tag{.465}$$

$$\frac{d^2}{dz_k dy_j} a_{ijkl} = y_{ji}z_{kj} - y_{ji}z_{lk} - y_{kj}z_{ji} + y_{kj}z_{lk} - y_{lk}z_{kj} - \left(y_j - y_l\right)\left(z_i - z_k\right) \tag{.466}$$

$$\frac{d^2}{dx_l dy_j} a_{ijkl} = x_{ji}y_{kj} - x_{kj}y_{ji} - y_{kj}\left(x_i - x_k\right) \tag{.467}$$

$$\frac{d^2}{dy_l dy_j} a_{ijkl} = x_{kj} \left(x_i - x_k \right) + z_{kj} \left(z_i - z_k \right) \tag{.468}$$

$$\frac{d^2}{dz_l dy_j} a_{ijkl} = -y_{ji} z_{kj} + y_{kj} z_{ji} - y_{kj} \left(z_i - z_k \right) \tag{.469}$$

$$\frac{d^2}{dz_j^2} a_{ijkl} = 2 x_{lk} \left(x_i - x_k \right) + 2 y_{lk} \left(y_i - y_k \right) \tag{.470}$$

$$\frac{d^2}{dx_k dz_j} a_{ijkl} = -x_{ji} z_{kj} + x_{kj} z_{ji} - x_{kj} z_{lk} - x_{lk} z_{ji} + x_{lk} z_{kj} + \left(x_i - x_k \right) \left(-z_j + z_l \right) \tag{.471}$$

$$\frac{d^2}{dy_k dz_j} a_{ijkl} = -y_{ji} z_{kj} + y_{kj} z_{ji} - y_{kj} z_{lk} - y_{lk} z_{ji} + y_{lk} z_{kj} + \left(y_i - y_k \right) \left(-z_j + z_l \right) \tag{.472}$$

$$\frac{d^2}{dz_k dz_j} a_{ijkl} = x_{ji} x_{lk} + y_{ji} y_{lk} + \left(x_i - x_k \right) \left(x_j - x_l \right) + \left(y_i - y_k \right) \left(y_j - y_l \right) \tag{.473}$$

$$\frac{d^2}{dx_l dz_j} a_{ijkl} = x_{ji} z_{kj} - x_{kj} z_{ji} - z_{kj} \left(x_i - x_k \right) \tag{.474}$$

$$\frac{d^2}{dy_l dz_j} a_{ijkl} = y_{ji} z_{kj} - y_{kj} z_{ji} - z_{kj} \left(y_i - y_k \right) \tag{.475}$$

$$\frac{d^2}{dz_l dz_j} a_{ijkl} = x_{kj} \left(x_i - x_k \right) + y_{kj} \left(y_i - y_k \right) \tag{.476}$$

$$\frac{d^2}{dx_k^2} a_{ijkl} = -y_{ji} \left(-y_j + y_l \right) + y_{ji} \left(y_j - y_l \right) - z_{ji} \left(-z_j + z_l \right) + z_{ji} \left(z_j - z_l \right) \tag{.477}$$

$$\frac{d^2}{dy_k dx_k} a_{ijkl} = -x_{ji} \left(y_j - y_l \right) - y_{ji} \left(x_j - x_l \right) \tag{.478}$$

$$\frac{d^2}{dz_k dx_k} a_{ijkl} = -x_{ji} \left(z_j - z_l \right) - z_{ji} \left(x_j - x_l \right) \tag{.479}$$

$$\frac{d^2}{dx_l dx_k} a_{ijkl} = y_{ji} y_{kj} + z_{ji} z_{kj} \tag{.480}$$

$$\frac{d^2}{dy_l dx_k} a_{ijkl} = x_{ji} y_{kj} - 2x_{kj} y_{ji} \tag{.481}$$

$$\frac{d^2}{dz_l dx_k} a_{ijkl} = x_{ji} z_{kj} - 2x_{kj} z_{ji} \tag{.482}$$

$$\frac{d^2}{dy_k^2} a_{ijkl} = 2x_{ji} \left(x_j - x_l \right) - z_{ji} \left(-z_j + z_l \right) + z_{ji} \left(z_j - z_l \right) \tag{.483}$$

$$\frac{d^2}{dz_k dy_k} a_{ijkl} = -y_{ji} \left(z_j - z_l \right) - z_{ji} \left(y_j - y_l \right) \tag{.484}$$

$$\frac{d^2}{dx_l dy_k} a_{ijkl} = -2x_{ji} y_{kj} + x_{kj} y_{ji} \tag{.485}$$

$$\frac{d^2}{dy_l dy_k} a_{ijkl} = x_{ji} x_{kj} + z_{ji} z_{kj} \tag{.486}$$

$$\frac{d^2}{dz_l dy_k} a_{ijkl} = y_{ji} z_{kj} - 2y_{kj} z_{ji} \tag{.487}$$

$$\frac{d^2}{dz_k^2} a_{ijkl} = 2x_{ji} \left(x_j - x_l \right) + 2y_{ji} \left(y_j - y_l \right) \tag{.488}$$

$$\frac{d^2}{dx_l dz_k} a_{ijkl} = -2x_{ji} z_{kj} + x_{kj} z_{ji} \tag{.489}$$

$$\frac{d^2}{dy_l dz_k} a_{ijkl} = -2y_{ji} z_{kj} + y_{kj} z_{ji} \tag{.490}$$

$$\frac{d^2}{dz_l dz_k} a_{ijkl} = x_{ji} x_{kj} + y_{ji} y_{kj} \tag{.491}$$

$$\frac{d^2}{dx_l^2} a_{ijkl} = 0 \tag{.492}$$

$$\frac{d^2}{dy_l dx_l} a_{ijkl} = 0 \tag{.493}$$

$$\frac{d^2}{dz_l dx_l} a_{ijkl} = 0 \tag{.494}$$

$$\frac{d^2}{dy_l^2} a_{ijkl} = 0 \tag{.495}$$

$$\frac{d^2}{dz_l dy_l} a_{ijkl} = 0 \tag{.496}$$

$$\frac{d^2}{dz_l^2} a_{ijkl} = 0 \tag{.497}$$

Derivatives of d_{ijkl}

The first and second order derivatives of d_{ijkl} with respect to the coordinates x, y and z of the participating atoms i,j,k and l are:

$$\frac{d}{dx_i} d_{ijkl} = r_{kj} \left(-y_{kj} z_{lk} + y_{lk} z_{kj} \right) \tag{.498}$$

$$\frac{d}{dy_i} d_{ijkl} = r_{kj} \left(x_{kj} z_{lk} - x_{lk} z_{kj} \right) \tag{.499}$$

$$\frac{d}{dz_i} d_{ijkl} = r_{kj} \left(-x_{kj} y_{lk} + x_{lk} y_{kj} \right) \tag{.500}$$

$$\frac{d}{dx_j} d_{ijkl} = r_{kj} \left(y_{ji} z_{lk} + y_{kj} z_{lk} - y_{lk} z_{ji} - y_{lk} z_{kj} \right) - \frac{x_{kj} \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}} \tag{.501}$$

$$\frac{d}{dy_j} d_{ijkl} = r_{kj} \left(-x_{ji} z_{lk} - x_{kj} z_{lk} + x_{lk} z_{ji} + x_{lk} z_{kj} \right) - \frac{y_{kj} \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}} \tag{.502}$$

$$\frac{d}{dz_j}d_{ijkl} = r_{kj}\left(x_{ji}y_{lk} + x_{kj}y_{lk} - x_{lk}y_{ji} - x_{lk}y_{kj}\right) - \frac{z_{kj}\left(x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + y_{ji}\left(-x_{kj}z_{lk} + x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)\right)}{r_{kj}} \tag{.503}$$

$$\frac{d}{dx_k}d_{ijkl} = r_{kj}\left(y_{ji}\left(z_j - z_l\right) + z_{ji}\left(-y_j + y_l\right)\right) + \frac{x_{kj}\left(x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + y_{ji}\left(-x_{kj}z_{lk} + x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)\right)}{r_{kj}} \tag{.504}$$

$$\frac{d}{dy_k}d_{ijkl} = r_{kj}\left(x_{ji}\left(-z_j + z_l\right) + z_{ji}\left(x_j - x_l\right)\right) + \frac{y_{kj}\left(x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + y_{ji}\left(-x_{kj}z_{lk} + x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)\right)}{r_{kj}} \tag{.505}$$

$$\frac{d}{dz_k}d_{ijkl} = r_{kj}\left(x_{ji}\left(y_j - y_l\right) + y_{ji}\left(-x_j + x_l\right)\right) + \frac{z_{kj}\left(x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + y_{ji}\left(-x_{kj}z_{lk} + x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)\right)}{r_{kj}} \tag{.506}$$

$$\frac{d}{dx_l}d_{ijkl} = r_{kj}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) \tag{.507}$$

$$\frac{d}{dy_l}d_{ijkl} = r_{kj}\left(-x_{ji}z_{kj} + x_{kj}z_{ji}\right) \tag{.508}$$

$$\frac{d}{dz_l}d_{ijkl} = r_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) \tag{.509}$$

$$\frac{d^2}{dx_i^2}d_{ijkl} = 0 \tag{.510}$$

$$\frac{d^2}{dy_i dx_i}d_{ijkl} = 0 \tag{.511}$$

$$\frac{d^2}{dz_i dx_i} d_{ijkl} = 0 \tag{.512}$$

$$\frac{d^2}{dx_j dx_i} d_{ijkl} = -\frac{x_{kj} \left(-y_{kj} z_{lk} + y_{lk} z_{kj}\right)}{r_{kj}} \tag{.513}$$

$$\frac{d^2}{dy_j dx_i} d_{ijkl} = r_{kj} z_{lk} - \frac{y_{kj} \left(-y_{kj} z_{lk} + y_{lk} z_{kj}\right)}{r_{kj}} \tag{.514}$$

$$\frac{d^2}{dz_j dx_i} d_{ijkl} = -r_{kj} y_{lk} - \frac{z_{kj} \left(-y_{kj} z_{lk} + y_{lk} z_{kj}\right)}{r_{kj}} \tag{.515}$$

$$\frac{d^2}{dx_k dx_i} d_{ijkl} = \frac{x_{kj} \left(-y_{kj} z_{lk} + y_{lk} z_{kj}\right)}{r_{kj}} \tag{.516}$$

$$\frac{d^2}{dy_k dx_i} d_{ijkl} = r_{kj} \left(z_j - z_l\right) + \frac{y_{kj} \left(-y_{kj} z_{lk} + y_{lk} z_{kj}\right)}{r_{kj}} \tag{.517}$$

$$\frac{d^2}{dz_k dx_i} d_{ijkl} = r_{kj} \left(-y_j + y_l\right) + \frac{z_{kj} \left(-y_{kj} z_{lk} + y_{lk} z_{kj}\right)}{r_{kj}} \tag{.518}$$

$$\frac{d^2}{dx_l dx_i} d_{ijkl} = 0 \tag{.519}$$

$$\frac{d^2}{dy_l dx_i} d_{ijkl} = r_{kj} z_{kj} \tag{.520}$$

$$\frac{d^2}{dz_l dx_i} d_{ijkl} = -r_{kj} y_{kj} \tag{.521}$$

$$\frac{d^2}{dy_i^2} d_{ijkl} = 0 \tag{.522}$$

$$\frac{d^2}{dz_i dy_i} d_{ijkl} = 0 \tag{.523}$$

$$\frac{d^2}{dx_j dy_i} d_{ijkl} = -r_{kj} z_{lk} - \frac{x_{kj} \left(x_{kj} z_{lk} - x_{lk} z_{kj} \right)}{r_{kj}} \tag{.524}$$

$$\frac{d^2}{dy_j dy_i} d_{ijkl} = -\frac{y_{kj} \left(x_{kj} z_{lk} - x_{lk} z_{kj} \right)}{r_{kj}} \tag{.525}$$

$$\frac{d^2}{dz_j dy_i} d_{ijkl} = r_{kj} x_{lk} - \frac{z_{kj} \left(x_{kj} z_{lk} - x_{lk} z_{kj} \right)}{r_{kj}} \tag{.526}$$

$$\frac{d^2}{dx_k dy_i} d_{ijkl} = r_{kj} \left(-z_j + z_l \right) + \frac{x_{kj} \left(x_{kj} z_{lk} - x_{lk} z_{kj} \right)}{r_{kj}} \tag{.527}$$

$$\frac{d^2}{dy_k dy_i} d_{ijkl} = \frac{y_{kj}\left(x_{kj} z_{lk} - x_{lk} z_{kj}\right)}{r_{kj}} \tag{.528}$$

$$\frac{d^2}{dz_k dy_i} d_{ijkl} = r_{kj}\left(x_j - x_l\right) + \frac{z_{kj}\left(x_{kj} z_{lk} - x_{lk} z_{kj}\right)}{r_{kj}} \tag{.529}$$

$$\frac{d^2}{dx_l dy_i} d_{ijkl} = -r_{kj} z_{kj} \tag{.530}$$

$$\frac{d^2}{dy_l dy_i} d_{ijkl} = 0 \tag{.531}$$

$$\frac{d^2}{dz_l dy_i} d_{ijkl} = r_{kj} x_{kj} \tag{.532}$$

$$\frac{d^2}{dz_i^2} d_{ijkl} = 0 \tag{.533}$$

$$\frac{d^2}{dx_j dz_i} d_{ijkl} = r_{kj} y_{lk} - \frac{x_{kj}\left(-x_{kj} y_{lk} + x_{lk} y_{kj}\right)}{r_{kj}} \tag{.534}$$

$$\frac{d^2}{dy_j dz_i} d_{ijkl} = -r_{kj} x_{lk} - \frac{y_{kj}\left(-x_{kj} y_{lk} + x_{lk} y_{kj}\right)}{r_{kj}} \tag{.535}$$

$$\frac{d^2}{dz_j dz_i} d_{ijkl} = -\frac{z_{kj}\left(-x_{kj}y_{lk} + x_{lk}y_{kj}\right)}{r_{kj}} \tag{.536}$$

$$\frac{d^2}{dx_k dz_i} d_{ijkl} = r_{kj}\left(y_j - y_l\right) + \frac{x_{kj}\left(-x_{kj}y_{lk} + x_{lk}y_{kj}\right)}{r_{kj}} \tag{.537}$$

$$\frac{d^2}{dy_k dz_i} d_{ijkl} = r_{kj}\left(-x_j + x_l\right) + \frac{y_{kj}\left(-x_{kj}y_{lk} + x_{lk}y_{kj}\right)}{r_{kj}} \tag{.538}$$

$$\frac{d^2}{dz_k dz_i} d_{ijkl} = \frac{z_{kj}\left(-x_{kj}y_{lk} + x_{lk}y_{kj}\right)}{r_{kj}} \tag{.539}$$

$$\frac{d^2}{dx_l dz_i} d_{ijkl} = r_{kj}y_{kj} \tag{.540}$$

$$\frac{d^2}{dy_l dz_i} d_{ijkl} = -r_{kj}x_{kj} \tag{.541}$$

$$\frac{d^2}{dz_l dz_i} d_{ijkl} = 0 \tag{.542}$$

$$\frac{d^2}{dx_j^2} d_{ijkl} = -\frac{2x_{kj}\left(y_{ji}z_{lk} + y_{kj}z_{lk} - y_{lk}z_{ji} - y_{lk}z_{kj}\right)}{r_{kj}} + \frac{x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + y_{ji}\left(-x_{kj}z_{lk} + x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)}{r_{kj}} - \frac{x_{kj}^2\left(x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + y_{ji}\left(-x_{kj}z_{lk} + x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)\right)}{r_{kj}^3} \tag{.543}$$

$$\frac{d^2}{dy_j dx_j} d_{ijkl} = -\frac{x_{kj}\left(-x_{ji}z_{lk} - x_{kj}z_{lk} + x_{lk}z_{ji} + x_{lk}z_{kj}\right)}{r_{kj}} - \frac{y_{kj}\left(y_{ji}z_{lk} + y_{kj}z_{lk} - y_{lk}z_{ji} - y_{lk}z_{kj}\right)}{r_{kj}} - \frac{x_{kj}y_{kj}\left(x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + y_{ji}\left(-x_{kj}z_{lk} + x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)\right)}{r_{kj}^3} \tag{.544}$$

$$\frac{d^2}{dz_j dx_j} d_{ijkl} = -\frac{x_{kj}\left(x_{ji}y_{lk} + x_{kj}y_{lk} - x_{lk}y_{ji} - x_{lk}y_{kj}\right)}{r_{kj}} - \frac{z_{kj}\left(y_{ji}z_{lk} + y_{kj}z_{lk} - y_{lk}z_{ji} - y_{lk}z_{kj}\right)}{r_{kj}} - \frac{x_{kj}z_{kj}\left(x_{ji}\left(y_{kj}z_{lk} - y_{lk}z_{kj}\right) + y_{ji}\left(-x_{kj}z_{lk} + x_{lk}z_{kj}\right) + z_{ji}\left(x_{kj}y_{lk} - x_{lk}y_{kj}\right)\right)}{r_{kj}^3} \tag{.545}$$

$$\frac{d^2}{dx_k dx_j} d_{ijkl} = -\frac{x_{kj}\left(y_{ji}\left(z_j - z_l\right) + z_{ji}\left(-y_j + y_l\right)\right)}{r_{kj}} + \frac{x_{kj}\left(y_{ji} z_{lk} + y_{kj} z_{lk} - y_{lk} z_{ji} - y_{lk} z_{kj}\right)}{r_{kj}} - \frac{x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)}{r_{kj}} + \frac{x_{kj}^2\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.546}$$

$$\frac{d^2}{dy_k dx_j} d_{ijkl} = r_{kj}\left(-z_i + z_l\right) - \frac{x_{kj}\left(x_{ji}\left(-z_j + z_l\right) + z_{ji}\left(x_j - x_l\right)\right)}{r_{kj}} + \frac{y_{kj}\left(y_{ji} z_{lk} + y_{kj} z_{lk} - y_{lk} z_{ji} - y_{lk} z_{kj}\right)}{r_{kj}} + \frac{x_{kj} y_{kj}\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.547}$$

$$\frac{d^2}{dz_k dx_j} d_{ijkl} = r_{kj}\left(y_i - y_l\right) - \frac{x_{kj}\left(x_{ji}\left(y_j - y_l\right) + y_{ji}\left(-x_j + x_l\right)\right)}{r_{kj}} + \frac{z_{kj}\left(y_{ji} z_{lk} + y_{kj} z_{lk} - y_{lk} z_{ji} - y_{lk} z_{kj}\right)}{r_{kj}} + \frac{x_{kj} z_{kj}\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.548}$$

$$\frac{d^2}{dx_l dx_j} d_{ijkl} = -\frac{x_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)}{r_{kj}} \tag{.549}$$

$$\frac{d^2}{dy_l dx_j} d_{ijkl} = r_{kj} \left(z_i - z_k \right) - \frac{x_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji} \right)}{r_{kj}} \tag{.550}$$

$$\frac{d^2}{dz_l dx_j} d_{ijkl} = r_{kj} \left(-y_i + y_k \right) - \frac{x_{kj} \left(x_{ji} y_{kj} - x_{kj} y_{ji} \right)}{r_{kj}} \tag{.551}$$

$$\begin{aligned} \frac{d^2}{dy_j^2} d_{ijkl} = & -\frac{2 y_{kj} \left(-x_{ji} z_{lk} - x_{kj} z_{lk} + x_{lk} z_{ji} + x_{lk} z_{kj} \right)}{r_{kj}} \\ & + \frac{x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right)}{r_{kj}} \\ & - \frac{y_{kj}^2 \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \end{aligned} \tag{.552}$$

$$\begin{aligned} \frac{d^2}{dz_j dy_j} d_{ijkl} = & -\frac{y_{kj} \left(x_{ji} y_{lk} + x_{kj} y_{lk} - x_{lk} y_{ji} - x_{lk} y_{kj} \right)}{r_{kj}} - \frac{z_{kj} \left(-x_{ji} z_{lk} - x_{kj} z_{lk} + x_{lk} z_{ji} + x_{lk} z_{kj} \right)}{r_{kj}} \\ & - \frac{y_{kj} z_{kj} \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \end{aligned} \tag{.553}$$

$$\begin{aligned} \frac{d^2}{dx_k dy_j} d_{ijkl} = & r_{kj} \left(z_i - z_l \right) + \frac{x_{kj} \left(-x_{ji} z_{lk} - x_{kj} z_{lk} + x_{lk} z_{ji} + x_{lk} z_{kj} \right)}{r_{kj}} \\ & - \frac{y_{kj} \left(y_{ji} \left(z_j - z_l \right) + z_{ji} \left(-y_j + y_l \right) \right)}{r_{kj}} \\ & + \frac{x_{kj} y_{kj} \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \end{aligned} \tag{.554}$$

$$\frac{d^2}{dy_k dy_j} d_{ijkl} = -\frac{y_{kj}\left(x_{ji}\left(-z_j + z_l\right) + z_{ji}\left(x_j - x_l\right)\right)}{r_{kj}} + \frac{y_{kj}\left(-x_{ji} z_{lk} - x_{kj} z_{lk} + x_{lk} z_{ji} + x_{lk} z_{kj}\right)}{r_{kj}} - \frac{x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)}{r_{kj}} + \frac{y_{kj}^2\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.555}$$

$$\frac{d^2}{dz_k dy_j} d_{ijkl} = r_{kj}\left(-x_i + x_l\right) - \frac{y_{kj}\left(x_{ji}\left(y_j - y_l\right) + y_{ji}\left(-x_j + x_l\right)\right)}{r_{kj}} + \frac{z_{kj}\left(-x_{ji} z_{lk} - x_{kj} z_{lk} + x_{lk} z_{ji} + x_{lk} z_{kj}\right)}{r_{kj}} + \frac{y_{kj} z_{kj}\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.556}$$

$$\frac{d^2}{dx_l dy_j} d_{ijkl} = r_{kj}\left(-z_i + z_k\right) - \frac{y_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)}{r_{kj}} \tag{.557}$$

$$\frac{d^2}{dy_l dy_j} d_{ijkl} = -\frac{y_{kj}\left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right)}{r_{kj}} \tag{.558}$$

$$\frac{d^2}{dz_l dy_j} d_{ijkl} = r_{kj}\left(x_i - x_k\right) - \frac{y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right)}{r_{kj}} \tag{.559}$$

$$\frac{d^2}{dz_j^2} d_{ijkl} = -\frac{2 z_{kj} \left(x_{ji} y_{lk} + x_{kj} y_{lk} - x_{lk} y_{ji} - x_{lk} y_{kj} \right)}{r_{kj}} + \frac{x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right)}{r_{kj}} - \frac{z_{kj}^2 \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \tag{.560}$$

$$\frac{d^2}{dx_k dz_j} d_{ijkl} = r_{kj} \left(-y_i + y_l \right) + \frac{x_{kj} \left(x_{ji} y_{lk} + x_{kj} y_{lk} - x_{lk} y_{ji} - x_{lk} y_{kj} \right)}{r_{kj}} - \frac{z_{kj} \left(y_{ji} \left(z_j - z_l \right) + z_{ji} \left(-y_j + y_l \right) \right)}{r_{kj}} + \frac{x_{kj} z_{kj} \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \tag{.561}$$

$$\frac{d^2}{dy_k dz_j} d_{ijkl} = r_{kj} \left(x_i - x_l \right) + \frac{y_{kj} \left(x_{ji} y_{lk} + x_{kj} y_{lk} - x_{lk} y_{ji} - x_{lk} y_{kj} \right)}{r_{kj}} - \frac{z_{kj} \left(x_{ji} \left(-z_j + z_l \right) + z_{ji} \left(x_j - x_l \right) \right)}{r_{kj}} + \frac{y_{kj} z_{kj} \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \tag{.562}$$

$$\frac{d^2}{dz_k dz_j} d_{ijkl} = -\frac{z_{kj} \left(x_{ji} \left(y_j - y_l \right) + y_{ji} \left(-x_j + x_l \right) \right)}{r_{kj}} + \frac{z_{kj} \left(x_{ji} y_{lk} + x_{kj} y_{lk} - x_{lk} y_{ji} - x_{lk} y_{kj} \right)}{r_{kj}} - \frac{x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right)}{r_{kj}} + \frac{z_{kj}^2 \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \tag{.563}$$

$$\frac{d^2}{dx_l dz_j} d_{ijkl} = r_{kj} \left(y_i - y_k \right) - \frac{z_{kj} \left(y_{ji} z_{kj} - y_{kj} z_{ji} \right)}{r_{kj}} \tag{.564}$$

$$\frac{d^2}{dy_l dz_j} d_{ijkl} = r_{kj} \left(-x_i + x_k \right) - \frac{z_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji} \right)}{r_{kj}} \tag{.565}$$

$$\frac{d^2}{dz_l dz_j} d_{ijkl} = -\frac{z_{kj} \left(x_{ji} y_{kj} - x_{kj} y_{ji} \right)}{r_{kj}} \tag{.566}$$

$$\begin{aligned} \frac{d^2}{dx_k^2} d_{ijkl} = & \frac{2 x_{kj} \left(y_{ji} \left(z_j - z_l \right) + z_{ji} \left(-y_j + y_l \right) \right)}{r_{kj}} \\ & + \frac{x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right)}{r_{kj}} \\ & - \frac{x_{kj}^2 \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \end{aligned} \tag{.567}$$

$$\begin{aligned} \frac{d^2}{dy_k dx_k} d_{ijkl} = & \frac{x_{kj} \left(x_{ji} \left(-z_j + z_l \right) + z_{ji} \left(x_j - x_l \right) \right)}{r_{kj}} \\ & + \frac{y_{kj} \left(y_{ji} \left(z_j - z_l \right) + z_{ji} \left(-y_j + y_l \right) \right)}{r_{kj}} \\ & - \frac{x_{kj} y_{kj} \left(x_{ji} \left(y_{kj} z_{lk} - y_{lk} z_{kj} \right) + y_{ji} \left(-x_{kj} z_{lk} + x_{lk} z_{kj} \right) + z_{ji} \left(x_{kj} y_{lk} - x_{lk} y_{kj} \right) \right)}{r_{kj}^3} \end{aligned} \tag{.568}$$

$$\frac{d^2}{dz_k dx_k} d_{ijkl} = \frac{x_{kj}\left(x_{ji}\left(y_j - y_l\right) + y_{ji}\left(-x_j + x_l\right)\right)}{r_{kj}} + \frac{z_{kj}\left(y_{ji}\left(z_j - z_l\right) + z_{ji}\left(-y_j + y_l\right)\right)}{r_{kj}} - \frac{x_{kj} z_{kj}\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.569}$$

$$\frac{d^2}{dx_l dx_k} d_{ijkl} = \frac{x_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)}{r_{kj}} \tag{.570}$$

$$\frac{d^2}{dy_l dx_k} d_{ijkl} = r_{kj} z_{ji} + \frac{x_{kj}\left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right)}{r_{kj}} \tag{.571}$$

$$\frac{d^2}{dz_l dx_k} d_{ijkl} = -r_{kj} y_{ji} + \frac{x_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right)}{r_{kj}} \tag{.572}$$

$$\frac{d^2}{dy_k^2} d_{ijkl} = \frac{2 y_{kj}\left(x_{ji}\left(-z_j + z_l\right) + z_{ji}\left(x_j - x_l\right)\right)}{r_{kj}} + \frac{x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)}{r_{kj}} - \frac{y_{kj}^2\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.573}$$

$$\frac{d^2}{dz_k dy_k} d_{ijkl} = \frac{y_{kj}\left(x_{ji}\left(y_j - y_l\right) + y_{ji}\left(-x_j + x_l\right)\right)}{r_{kj}} + \frac{z_{kj}\left(x_{ji}\left(-z_j + z_l\right) + z_{ji}\left(x_j - x_l\right)\right)}{r_{kj}} - \frac{y_{kj} z_{kj}\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.574}$$

$$\frac{d^2}{dx_l dy_k} d_{ijkl} = -r_{kj} z_{ji} + \frac{y_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)}{r_{kj}} \tag{.575}$$

$$\frac{d^2}{dy_l dy_k} d_{ijkl} = \frac{y_{kj}\left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right)}{r_{kj}} \tag{.576}$$

$$\frac{d^2}{dz_l dy_k} d_{ijkl} = r_{kj} x_{ji} + \frac{y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right)}{r_{kj}} \tag{.577}$$

$$\frac{d^2}{dz_k^2} d_{ijkl} = \frac{2 z_{kj}\left(x_{ji}\left(y_j - y_l\right) + y_{ji}\left(-x_j + x_l\right)\right)}{r_{kj}} + \frac{x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)}{r_{kj}} - \frac{z_{kj}^2\left(x_{ji}\left(y_{kj} z_{lk} - y_{lk} z_{kj}\right) + y_{ji}\left(-x_{kj} z_{lk} + x_{lk} z_{kj}\right) + z_{ji}\left(x_{kj} y_{lk} - x_{lk} y_{kj}\right)\right)}{r_{kj}^3} \tag{.578}$$

$$\frac{d^2}{dx_l dz_k} d_{ijkl} = r_{kj} y_{ji} + \frac{z_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)}{r_{kj}} \tag{.579}$$

$$\frac{d^2}{dy_l dz_k} d_{ijkl} = -r_{kj} x_{ji} + \frac{z_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji} \right)}{r_{kj}} \tag{.580}$$

$$\frac{d^2}{dz_l dz_k} d_{ijkl} = \frac{z_{kj} \left(x_{ji} y_{kj} - x_{kj} y_{ji} \right)}{r_{kj}} \tag{.581}$$

$$\frac{d^2}{dx_l^2} d_{ijkl} = 0 \tag{.582}$$

$$\frac{d^2}{dy_l dx_l} d_{ijkl} = 0 \tag{.583}$$

$$\frac{d^2}{dz_l dx_l} d_{ijkl} = 0 \tag{.584}$$

$$\frac{d^2}{dy_l^2} d_{ijkl} = 0 \tag{.585}$$

$$\frac{d^2}{dz_l dy_l} d_{ijkl} = 0 \tag{.586}$$

$$\frac{d^2}{dz_l^2} d_{ijkl} = 0 \tag{.587}$$

Derivatives of b_{ijk}

The first and second order derivatives of b_{ijk} with respect to the coordinates x, y and z of the participating atoms i, j and k are:

$$\frac{d}{dx_i}b_{ijk} = \frac{-y_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) + z_{kj}\left(-x_{ji}z_{kj} + x_{kj}z_{ji}\right)}{b_{ijk}} \tag{.588}$$

$$\frac{d}{dy_i}b_{ijk} = \frac{x_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - z_{kj}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)}{b_{ijk}} \tag{.589}$$

$$\frac{d}{dz_i}b_{ijk} = \frac{-x_{kj}\left(-x_{ji}z_{kj} + x_{kj}z_{ji}\right) + y_{kj}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)}{b_{ijk}} \tag{.590}$$

$$\frac{d}{dx_j}b_{ijk} = \frac{y_{ji}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) + y_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - z_{ji}\left(-x_{ji}z_{kj} + x_{kj}z_{ji}\right) - z_{kj}\left(-x_{ji}z_{kj} + x_{kj}z_{ji}\right)}{b_{ijk}} \tag{.591}$$

$$\frac{d}{dy_j}b_{ijk} = \frac{-x_{ji}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - x_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) + z_{ji}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) + z_{kj}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)}{b_{ijk}} \tag{.592}$$

$$\frac{d}{dz_j}b_{ijk} = \frac{x_{ji}\left(-x_{ji}z_{kj} + x_{kj}z_{ji}\right) + x_{kj}\left(-x_{ji}z_{kj} + x_{kj}z_{ji}\right) - y_{ji}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right) - y_{kj}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)}{b_{ijk}} \tag{.593}$$

$$\frac{d}{dx_k}b_{ijk} = \frac{-y_{ji}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - z_{ji}\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right)}{b_{ijk}} \tag{.594}$$

$$\frac{d}{dy_k}b_{ijk} = \frac{-x_{ji}\left(-x_{ji}y_{kj} + x_{kj}y_{ji}\right) - z_{ji}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)}{b_{ijk}} \tag{.595}$$

$$\frac{d}{dz_k}b_{ijk} = \frac{-x_{ji}\left(-x_{ji}z_{kj} + x_{kj}z_{ji}\right) - y_{ji}\left(-y_{ji}z_{kj} + y_{kj}z_{ji}\right)}{b_{ijk}} \tag{.596}$$

$$\frac{d^2}{dx_i dx_i}b_{ijk} = \frac{b_{ijk}^2\left(y_{kj}^2 + z_{kj}^2\right) - \left(y_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) + z_{kj}\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.597}$$

$$\frac{d^2}{dy_i dx_i}b_{ijk} = \frac{-b_{ijk}^2 x_{kj}y_{kj} + \left(x_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) - z_{kj}\left(y_{ji}z_{kj} - y_{kj}z_{ji}\right)\right)\left(y_{kj}\left(x_{ji}y_{kj} - x_{kj}y_{ji}\right) + z_{kj}\left(x_{ji}z_{kj} - x_{kj}z_{ji}\right)\right)}{b_{ijk}^3} \tag{.598}$$

$$\frac{d^2}{dz_i dx_i} b_{ijk} = \frac{-b_{ijk}^2 x_{kj} z_{kj} + \left(x_{kj}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right) - y_{kj}\left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right)\left(y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{kj}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.599}$$

$$\frac{d^2}{dx_j dx_i} b_{ijk} = \frac{b_{ijk}^2\left(-y_{kj}\left(y_{ji} + y_{kj}\right) - z_{kj}\left(z_{ji} + z_{kj}\right)\right) + \left(y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{kj}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)\left(\left(y_{ji} + y_{kj}\right)\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + \left(z_{ji} + z_{kj}\right)\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.600}$$

$$\frac{d^2}{dy_j dx_i} b_{ijk} = \frac{-b_{ijk}^2\left(-x_{ji} y_{kj} + x_{kj} y_{ji} - y_{kj}\left(x_{ji} + x_{kj}\right)\right) + \left(y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{kj}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)\left(\left(-x_{ji} - x_{kj}\right)\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + \left(z_{ji} + z_{kj}\right)\left(y_{ji} z_{kj} - y_{kj}\right.\right.}{b_{ijk}^3} \tag{.601}$$

$$\frac{d^2}{dz_j dx_i} b_{ijk} = \frac{-b_{ijk}^2\left(-x_{ji} z_{kj} + x_{kj} z_{ji} - z_{kj}\left(x_{ji} + x_{kj}\right)\right) + \left(y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{kj}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)\left(\left(-x_{ji} - x_{kj}\right)\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right) + \left(y_{ji} + y_{kj}\right)\left(-y_{ji} z_{kj} + y\right.\right.}{b_{ijk}^3} \tag{.602}$$

$$\frac{d^2}{dx_k dx_i} b_{ijk} = \frac{b_{ijk}^2\left(y_{ji} y_{kj} + z_{ji} z_{kj}\right) - \left(y_{ji}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{ji}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)\left(y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{kj}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.603}$$

$$\frac{d^2}{dy_k dx_i} b_{ijk} = \frac{b_{ijk}^2\left(-2 x_{ji} y_{kj} + x_{kj} y_{ji}\right) + \left(x_{ji}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) - z_{ji}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)\left(y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{kj}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.604}$$

$$\frac{d^2}{dz_k dx_i} b_{ijk} = \frac{b_{ijk}^2\left(-2 x_{ji} z_{kj} + x_{kj} z_{ji}\right) + \left(x_{ji}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right) - y_{ji}\left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right)\left(y_{kj}\left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{kj}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.605}$$

$$\frac{d^2}{dy_i dy_i} b_{ijk} = \frac{b_{ijk}^2\left(x_{kj}^2 + z_{kj}^2\right) - \left(x_{kj}\left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.606}$$

$$\frac{d^2}{dz_i dy_i} b_{ijk} = \frac{-b_{ijk}^2 y_{kj} z_{kj} + \left(x_{kj}\left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)\left(-x_{kj}\left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.607}$$

$$\frac{d^2}{dx_j dy_i} b_{ijk} = \frac{-b_{ijk}^2\left(x_{ji} y_{kj} - x_{kj} y_{ji} - x_{kj}\left(y_{ji} + y_{kj}\right)\right) + \left(x_{kj}\left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)\left(\left(-y_{ji} - y_{kj}\right)\left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + \left(z_{ji} + z_{kj}\right)\left(x_{ji} z_{kj} - x\right.\right.}{b_{ijk}^3} \tag{.608}$$

$$\frac{d^2}{dy_j dy_i} b_{ijk} = \frac{b_{ijk}^2\left(-x_{kj}\left(x_{ji} + x_{kj}\right) - z_{kj}\left(z_{ji} + z_{kj}\right)\right) + \left(x_{kj}\left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{kj}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)\left(\left(x_{ji} + x_{kj}\right)\left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + \left(z_{ji} + z_{kj}\right)\left(y_{ji} z_{kj} - y_{kj} z\right.\right.}{b_{ijk}^3} \tag{.609}$$

$$\frac{d^2}{dz_j dy_i} b_{ijk} = \frac{-b_{ijk}^2 \left(-y_{ji} z_{kj} + y_{kj} z_{ji} - z_{kj} \left(y_{ji} + y_{kj}\right)\right) + \left(x_{kj} \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{kj} \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right) \left(\left(x_{ji} + x_{kj}\right) \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + \left(-y_{ji} - y_{kj}\right) \left(y_{ji} z_{kj} - \right.\right.}{b_{ijk}^3} \tag{.610}$$

$$\frac{d^2}{dx_k dy_i} b_{ijk} = \frac{b_{ijk}^2 \left(x_{ji} y_{kj} - 2 x_{kj} y_{ji}\right) + \left(x_{kj} \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{kj} \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right) \left(y_{ji} \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) - z_{ji} \left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.611}$$

$$\frac{d^2}{dy_k dy_i} b_{ijk} = \frac{b_{ijk}^2 \left(x_{ji} x_{kj} + z_{ji} z_{kj}\right) - \left(x_{ji} \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{ji} \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right) \left(x_{kj} \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{kj} \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.612}$$

$$\frac{d^2}{dz_k dy_i} b_{ijk} = \frac{b_{ijk}^2 \left(-2 y_{ji} z_{kj} + y_{kj} z_{ji}\right) + \left(-x_{ji} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{ji} \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right) \left(x_{kj} \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{kj} \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.613}$$

$$\frac{d^2}{dz_i dz_i} b_{ijk} = \frac{b_{ijk}^2 \left(x_{kj}^2 + y_{kj}^2\right) - \left(x_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{kj} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.614}$$

$$\frac{d^2}{dx_j dz_i} b_{ijk} = \frac{-b_{ijk}^2 \left(x_{ji} z_{kj} - x_{kj} z_{ji} - x_{kj} \left(z_{ji} + z_{kj}\right)\right) + \left(x_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{kj} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right) \left(\left(y_{ji} + y_{kj}\right) \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + \left(-z_{ji} - z_{kj}\right) \left(-x_{ji} z_{kj} + \right.\right.}{b_{ijk}^3} \tag{.615}$$

$$\frac{d^2}{dy_j dz_i} b_{ijk} = \frac{-b_{ijk}^2 \left(y_{ji} z_{kj} - y_{kj} z_{ji} - y_{kj} \left(z_{ji} + z_{kj}\right)\right) + \left(x_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{kj} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right) \left(\left(x_{ji} + x_{kj}\right) \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + \left(-z_{ji} - z_{kj}\right) \left(-y_{ji} z_{kj} - \right.\right.}{b_{ijk}^3} \tag{.616}$$

$$\frac{d^2}{dz_j dz_i} b_{ijk} = \frac{b_{ijk}^2 \left(-x_{kj} \left(x_{ji} + x_{kj}\right) - y_{kj} \left(y_{ji} + y_{kj}\right)\right) + \left(x_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{kj} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right) \left(\left(x_{ji} + x_{kj}\right) \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + \left(y_{ji} + y_{kj}\right) \left(-y_{ji} z_{kj} + \right.\right.}{b_{ijk}^3} \tag{.617}$$

$$\frac{d^2}{dx_k dz_i} b_{ijk} = \frac{b_{ijk}^2 \left(x_{ji} z_{kj} - 2 x_{kj} z_{ji}\right) + \left(x_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{kj} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right) \left(-y_{ji} \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{ji} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.618}$$

$$\frac{d^2}{dy_k dz_i} b_{ijk} = \frac{b_{ijk}^2 \left(y_{ji} z_{kj} - 2 y_{kj} z_{ji}\right) + \left(-x_{ji} \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + z_{ji} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right) \left(x_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{kj} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.619}$$

$$\frac{d^2}{dz_k dz_i} b_{ijk} = \frac{b_{ijk}^2 \left(x_{ji} x_{kj} + y_{ji} y_{kj}\right) - \left(x_{ji} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{ji} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right) \left(x_{kj} \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + y_{kj} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.620}$$

$$\frac{d^2}{dx_j dx_j} b_{ijk} = \frac{b_{ijk}^2 \left(\left(y_{ji} + y_{kj}\right)^2 + \left(z_{ji} + z_{kj}\right)^2\right) - \left(\left(y_{ji} + y_{kj}\right) \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + \left(z_{ji} + z_{kj}\right) \left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.621}$$

$$\frac{d^2}{dy_j dx_j} b_{ijk} = \frac{b_{ijk}^2 \left(-x_{ji} - x_{kj}\right) \left(y_{ji} + y_{kj}\right) + \left(\left(-x_{ji} - x_{kj}\right) \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + \left(z_{ji} + z_{kj}\right) \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right) \left(\left(-y_{ji} - y_{kj}\right) \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) - \left(z_{ji} + z_{kj}\right) \left(x_{ji} z_{kj} -\right.}{b_{ijk}^3} \tag{.622}$$

$$\frac{d^2}{dz_j dx_j} b_{ijk} = \frac{b_{ijk}^2 \left(-x_{ji} - x_{kj}\right) \left(z_{ji} + z_{kj}\right) + \left(\left(-x_{ji} - x_{kj}\right) \left(x_{ji} z_{kj} - x_{kj} z_{ji}\right) + \left(y_{ji} + y_{kj}\right) \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right) \left(-\left(y_{ji} + y_{kj}\right) \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + \left(-z_{ji} - z_{kj}\right) \left(x_{ji} z_{k}\right.}{b_{ijk}^3} \tag{.623}$$

$$\frac{d^2}{dx_k dx_j} b_{ijk} = \frac{b_{ijk}^2 \left(-y_{ji} \left(y_{ji} + y_{kj}\right) - z_{ji} \left(z_{ji} + z_{kj}\right)\right) + \left(y_{ji} \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + z_{ji} \left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right) \left(\left(y_{ji} + y_{kj}\right) \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + \left(z_{ji} + z_{kj}\right) \left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right)}{b_{ijk}^3} \tag{.624}$$

$$\frac{d^2}{dy_k dx_j} b_{ijk} = \frac{-b_{ijk}^2 \left(-x_{ji} y_{kj} - x_{ji} \left(y_{ji} + y_{kj}\right) + x_{kj} y_{ji}\right) + \left(x_{ji} \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) - z_{ji} \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right) \left(\left(-y_{ji} - y_{kj}\right) \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) - \left(z_{ji} + z_{kj}\right) \left(x_{ji} z_{kj} - x_{kj} z\right.}{b_{ijk}^3} \tag{.625}$$

$$\frac{d^2}{dz_k dx_j} b_{ijk} = \frac{-b_{ijk}^2 \left(-x_{ji} z_{kj} - x_{ji} \left(z_{ji} + z_{kj}\right) + x_{kj} z_{ji}\right) + \left(x_{ji} \left(x_{ji} z_{kj} - x_{kj} z_{ji}\right) - y_{ji} \left(-y_{ji} z_{kj} + y_{kj} z_{ji}\right)\right) \left(-\left(y_{ji} + y_{kj}\right) \left(x_{ji} y_{kj} - x_{kj} y_{ji}\right) + \left(-z_{ji} - z_{kj}\right) \left(x_{ji} z_{kj} -\right.}{b_{ijk}^3} \tag{.626}$$

$$\frac{d^2}{dy_j dy_j} b_{ijk} = \frac{b_{ijk}^2 \left(\left(x_{ji} + x_{kj}\right)^2 + \left(z_{ji} + z_{kj}\right)^2\right) - \left(\left(x_{ji} + x_{kj}\right) \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + \left(z_{ji} + z_{kj}\right) \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.627}$$

$$\frac{d^2}{dz_j dy_j} b_{ijk} = \frac{b_{ijk}^2 \left(-y_{ji} - y_{kj}\right) \left(z_{ji} + z_{kj}\right) + \left(-\left(x_{ji} + x_{kj}\right) \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) + \left(-z_{ji} - z_{kj}\right) \left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right) \left(\left(x_{ji} + x_{kj}\right) \left(-x_{ji} z_{kj} + x_{kj} z_{ji}\right) + \left(-y_{ji} - y_{kj}\right) \left(y_{ji}\right.}{b_{ijk}^3} \tag{.628}$$

$$\frac{d^2}{dx_k dy_j} b_{ijk} = \frac{-b_{ijk}^2 \left(x_{ji} y_{kj} - x_{kj} y_{ji} - y_{ji} \left(x_{ji} + x_{kj}\right)\right) + \left(y_{ji} \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) - z_{ji} \left(x_{ji} z_{kj} - x_{kj} z_{ji}\right)\right) \left(\left(-x_{ji} - x_{kj}\right) \left(-x_{ji} y_{kj} + x_{kj} y_{ji}\right) - \left(z_{ji} + z_{kj}\right) \left(y_{ji} z_{kj} - y_{j}\right.}{b_{ijk}^3} \tag{.629}$$

$$\frac{d^2}{dy_k dy_j} b_{ijk} = \frac{b_{ijk}^2 \left(-x_{ji}\left(x_{ji}+x_{kj}\right) - z_{ji}\left(z_{ji}+z_{kj}\right)\right) + \left(x_{ji}\left(-x_{ji}y_{kj}+x_{kj}y_{ji}\right) + z_{ji}\left(y_{ji}z_{kj}-y_{kj}z_{ji}\right)\right)\left(\left(x_{ji}+x_{kj}\right)\left(-x_{ji}y_{kj}+x_{kj}y_{ji}\right) + \left(z_{ji}+z_{kj}\right)\left(y_{ji}z_{kj}-y_{kj}z_{j}\right.\right.}{b_{ijk}^3} \tag{.630}$$

$$\frac{d^2}{dz_k dy_j} b_{ijk} = \frac{-b_{ijk}^2 \left(-y_{ji}z_{kj} - y_{ji}\left(z_{ji}+z_{kj}\right) + y_{kj}z_{ji}\right) + \left(-x_{ji}\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right) + y_{ji}\left(y_{ji}z_{kj}-y_{kj}z_{ji}\right)\right)\left(-\left(x_{ji}+x_{kj}\right)\left(-x_{ji}y_{kj}+x_{kj}y_{ji}\right) + \left(-z_{ji}-z_{kj}\right)\left(y_{ji}z_{kj}\right.\right.}{b_{ijk}^3} \tag{.631}$$

$$\frac{d^2}{dz_j dz_j} b_{ijk} = \frac{b_{ijk}^2 \left(\left(x_{ji}+x_{kj}\right)^2 + \left(y_{ji}+y_{kj}\right)^2\right) - \left(\left(x_{ji}+x_{kj}\right)\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right) + \left(y_{ji}+y_{kj}\right)\left(-y_{ji}z_{kj}+y_{kj}z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.632}$$

$$\frac{d^2}{dx_k dz_j} b_{ijk} = \frac{-b_{ijk}^2 \left(x_{ji}z_{kj} - x_{kj}z_{ji} - z_{ji}\left(x_{ji}+x_{kj}\right)\right) + \left(-y_{ji}\left(x_{ji}y_{kj}-x_{kj}y_{ji}\right) + z_{ji}\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right)\right)\left(\left(-x_{ji}-x_{kj}\right)\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right) - \left(y_{ji}+y_{kj}\right)\left(-y_{ji}z_{kj}\right.\right.}{b_{ijk}^3} \tag{.633}$$

$$\frac{d^2}{dy_k dz_j} b_{ijk} = \frac{-b_{ijk}^2 \left(y_{ji}z_{kj} - y_{kj}z_{ji} - z_{ji}\left(y_{ji}+y_{kj}\right)\right) + \left(-x_{ji}\left(-x_{ji}y_{kj}+x_{kj}y_{ji}\right) + z_{ji}\left(-y_{ji}z_{kj}+y_{kj}z_{ji}\right)\right)\left(-\left(x_{ji}+x_{kj}\right)\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right) + \left(-y_{ji}-y_{kj}\right)\left(-y_{ji}z\right.\right.}{b_{ijk}^3} \tag{.634}$$

$$\frac{d^2}{dz_k dz_j} b_{ijk} = \frac{b_{ijk}^2 \left(-x_{ji}\left(x_{ji}+x_{kj}\right) - y_{ji}\left(y_{ji}+y_{kj}\right)\right) + \left(x_{ji}\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right) + y_{ji}\left(-y_{ji}z_{kj}+y_{kj}z_{ji}\right)\right)\left(\left(x_{ji}+x_{kj}\right)\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right) + \left(y_{ji}+y_{kj}\right)\left(-y_{ji}z_{kj}+y\right.\right.}{b_{ijk}^3} \tag{.635}$$

$$\frac{d^2}{dx_k dx_k} b_{ijk} = \frac{b_{ijk}^2 \left(y_{ji}^2 + z_{ji}^2\right) - \left(y_{ji}\left(-x_{ji}y_{kj}+x_{kj}y_{ji}\right) + z_{ji}\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.636}$$

$$\frac{d^2}{dy_k dx_k} b_{ijk} = \frac{-b_{ijk}^2 x_{ji}y_{ji} + \left(x_{ji}\left(-x_{ji}y_{kj}+x_{kj}y_{ji}\right) - z_{ji}\left(-y_{ji}z_{kj}+y_{kj}z_{ji}\right)\right)\left(y_{ji}\left(-x_{ji}y_{kj}+x_{kj}y_{ji}\right) + z_{ji}\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right)\right)}{b_{ijk}^3} \tag{.637}$$

$$\frac{d^2}{dz_k dx_k} b_{ijk} = \frac{-b_{ijk}^2 x_{ji}z_{ji} + \left(x_{ji}\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right) - y_{ji}\left(y_{ji}z_{kj}-y_{kj}z_{ji}\right)\right)\left(y_{ji}\left(-x_{ji}y_{kj}+x_{kj}y_{ji}\right) + z_{ji}\left(-x_{ji}z_{kj}+x_{kj}z_{ji}\right)\right)}{b_{ijk}^3} \tag{.638}$$

$$\frac{d^2}{dy_k dy_k} b_{ijk} = \frac{b_{ijk}^2 \left(x_{ji}^2 + z_{ji}^2\right) - \left(x_{ji}\left(x_{ji}y_{kj}-x_{kj}y_{ji}\right) + z_{ji}\left(-y_{ji}z_{kj}+y_{kj}z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.639}$$

$$\frac{d^2}{dz_k dy_k} b_{ijk} = \frac{-b_{ijk}^2 y_{ji}z_{ji} + \left(x_{ji}\left(x_{ji}y_{kj}-x_{kj}y_{ji}\right) + z_{ji}\left(-y_{ji}z_{kj}+y_{kj}z_{ji}\right)\right)\left(-x_{ji}\left(x_{ji}z_{kj}-x_{kj}z_{ji}\right) + y_{ji}\left(-y_{ji}z_{kj}+y_{kj}z_{ji}\right)\right)}{b_{ijk}^3} \tag{.640}$$

$$\frac{d^2}{dz_k dz_k} b_{ijk} = \frac{b_{ijk}^2 \left(x_{ji}^2 + y_{ji}^2\right) - \left(x_{ji}\left(x_{ji} z_{kj} - x_{kj} z_{ji}\right) + y_{ji}\left(y_{ji} z_{kj} - y_{kj} z_{ji}\right)\right)^2}{b_{ijk}^3} \tag{.641}$$

The first and second order derivatives of b_{ijk} are equivalent to the derivatives of b_{jkl} by interchanging atomic indices.

Coulomb Repulsion

A single Coulomb repulsion between two atoms i and j is given by

$$E_{\text{Coulomb},ij} = \frac{Q}{r_{ij}} \tag{.642}$$

Derivatives of $E_{\text{Coulomb},ij}$

The first and second order derivatives of $E_{\text{Coulomb},ij}$ with respect to the coordinates of the participating atoms are:

$$\frac{d}{dx_i} E_{\text{Coulomb},ij} = -\frac{Q \frac{d}{dx_i} r_{ij}}{r_{ij}^2} \tag{.643}$$

$$\frac{d}{dy_i} E_{\text{Coulomb},ij} = -\frac{Q \frac{d}{dy_i} r_{ij}}{r_{ij}^2} \tag{.644}$$

$$\frac{d}{dz_i} E_{\text{Coulomb},ij} = -\frac{Q \frac{d}{dz_i} r_{ij}}{r_{ij}^2} \tag{.645}$$

$$\frac{d}{dx_j} E_{\text{Coulomb},ij} = -\frac{Q \frac{d}{dx_j} r_{ij}}{r_{ij}^2} \tag{.646}$$

$$\frac{d}{dy_j} E_{\text{Coulomb},ij} = -\frac{Q \frac{d}{dy_j} r_{ij}}{r_{ij}^2} \tag{.647}$$

$$\frac{d}{dz_j}E_{\text{Coulomb},ij} = -\frac{Q\frac{d}{dz_j}r_{ij}}{r_{ij}^2} \tag{.648}$$

$$\frac{d^2}{dx_i^2}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dx_i^2}r_{ij}}{r_{ij}^2} + \frac{2Q\left(\frac{d}{dx_i}r_{ij}\right)^2}{r_{ij}^3} \tag{.649}$$

$$\frac{d^2}{dy_i dx_i}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dy_i dx_i}r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_i}r_{ij}\frac{d}{dy_i}r_{ij}}{r_{ij}^3} \tag{.650}$$

$$\frac{d^2}{dz_i dx_i}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_i dx_i}r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_i}r_{ij}\frac{d}{dz_i}r_{ij}}{r_{ij}^3} \tag{.651}$$

$$\frac{d^2}{dx_j dx_i}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dx_j dx_i}r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_i}r_{ij}\frac{d}{dx_j}r_{ij}}{r_{ij}^3} \tag{.652}$$

$$\frac{d^2}{dy_j dx_i}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dy_j dx_i}r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_i}r_{ij}\frac{d}{dy_j}r_{ij}}{r_{ij}^3} \tag{.653}$$

$$\frac{d^2}{dz_j dx_i}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_j dx_i}r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_i}r_{ij}\frac{d}{dz_j}r_{ij}}{r_{ij}^3} \tag{.654}$$

$$\frac{d^2}{dy_i^2}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dy_i^2}r_{ij}}{r_{ij}^2} + \frac{2Q\left(\frac{d}{dy_i}r_{ij}\right)^2}{r_{ij}^3} \tag{.655}$$

$$\frac{d^2}{dz_i dy_i}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_i dy_i}r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dy_i}r_{ij}\frac{d}{dz_i}r_{ij}}{r_{ij}^3} \tag{.656}$$

$$\frac{d^2}{dy_i dx_j}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dy_i dx_j}r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_j}r_{ij}\frac{d}{dy_i}r_{ij}}{r_{ij}^3} \tag{.657}$$

$$\frac{d^2}{dy_j dy_i}E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dy_j dy_i}r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dy_i}r_{ij}\frac{d}{dy_j}r_{ij}}{r_{ij}^3} \tag{.658}$$

$$\frac{d^2}{dz_j dy_i} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_j dy_i} r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dy_i} r_{ij} \frac{d}{dz_j} r_{ij}}{r_{ij}^3} \tag{.659}$$

$$\frac{d^2}{dz_i^2} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_i^2} r_{ij}}{r_{ij}^2} + \frac{2Q\left(\frac{d}{dz_i} r_{ij}\right)^2}{r_{ij}^3} \tag{.660}$$

$$\frac{d^2}{dx_j dz_i} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_i dx_j} r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_j} r_{ij} \frac{d}{dz_i} r_{ij}}{r_{ij}^3} \tag{.661}$$

$$\frac{d^2}{dy_j dz_i} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_i dy_j} r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dy_j} r_{ij} \frac{d}{dz_i} r_{ij}}{r_{ij}^3} \tag{.662}$$

$$\frac{d^2}{dz_j dz_i} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_j dz_i} r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dz_i} r_{ij} \frac{d}{dz_j} r_{ij}}{r_{ij}^3} \tag{.663}$$

$$\frac{d^2}{dx_j^2} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dx_j^2} r_{ij}}{r_{ij}^2} + \frac{2Q\left(\frac{d}{dx_j} r_{ij}\right)^2}{r_{ij}^3} \tag{.664}$$

$$\frac{d^2}{dy_j dx_j} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dy_j dx_j} r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_j} r_{ij} \frac{d}{dy_j} r_{ij}}{r_{ij}^3} \tag{.665}$$

$$\frac{d^2}{dz_j dx_j} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_j dx_j} r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dx_j} r_{ij} \frac{d}{dz_j} r_{ij}}{r_{ij}^3} \tag{.666}$$

$$\frac{d^2}{dy_j^2} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dy_j^2} r_{ij}}{r_{ij}^2} + \frac{2Q\left(\frac{d}{dy_j} r_{ij}\right)^2}{r_{ij}^3} \tag{.667}$$

$$\frac{d^2}{dz_j dy_j} E_{\text{Coulomb},ij} = -\frac{Q\frac{d^2}{dz_j dy_j} r_{ij}}{r_{ij}^2} + \frac{2Q\frac{d}{dy_j} r_{ij} \frac{d}{dz_j} r_{ij}}{r_{ij}^3} \tag{.668}$$