Incorporating Quantum Correlation in Extended Molecular Systems Using Advanced Electronic Structure Methods in Quantum Chemistry

by

Mostafa Javaheri Moghadam

Bachelor of Science, University of Environment, 2011
Master of Science, Golestan University (GU), 2015

A Dissertation Submitted in Partial Fulfilment of
the Requirements for the Degree of

Doctor of Philosophy

In the Graduate Academic Unit of Chemistry

Supervisor: Stijn De Baerdemacker, Ph.D., Dept. of Chemistry

Examining Board: Yang Qu, Ph.D., Dept. of Chemistry
Dion G. Durnford, Ph.D., Dept. of Biology
Gavin Douglas, Ph.D., Dept. of Biology

External Examiner: Chérif F. Matta, Ph.D., Dept. of Chemistry & Physics,
Mount Saint Vincent University

This dissertation is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

April, 2026

© Mostafa Javaheri Moghadam, 2026

Abstract

Understanding electronic correlations in biomolecules is critical for advancing quantum chemistry, drug design, and molecular biology. This thesis presents a quantum information theory approach to analyzing electronic correlations in biomolecular systems using Mutual Information (MI), a measure derived from Reduced Density Matrices (RDMs) and orbital entropies. We introduce Atomic Mutual Information (AMI) as a novel metric that quantifies total correlation between atoms by summing MI contributions from all orbital pairs connecting them, enabling detailed characterization of both intra- and inter-molecular interactions.

The methodology is first applied to small systems, including dipeptides. To ensure accuracy, density matrix renormalization group (DMRG) calculations serve as a high-level benchmark against which less computationally demanding methods such as Density Functional Theory (DFT) and restricted Hartree-Fock (RHF) are evaluated. We then extend our approach to larger peptides by introducing Fragment-wise Mutual Information (FMI), a coarse-grained version of AMI that enables inter-fragment correlation analysis. To study dynamic behavior, we analyze the evolution of FMI by extracting snapshots from molecular dynamics (MD) trajectories. The results demonstrate that increases in FMI correlate with folding and stabilization of peptide structures, offering a quantum-informed descriptor of conformational changes that complements classical metrics. Finally, the method is scaled up to small proteins through a cut-wise reconstruction strategy. By stitching together results from over-

lapping fragments, we approximate full-system FMI with reasonable accuracy. Validation on the insulin protein confirms the method's ability to capture key features, including hydrogen bonding, secondary structures, and disulfide bridges.

This work introduces scalable and interpretable tools for mapping electronic correlations in biomolecular systems, bridging quantum chemistry and molecular modeling, with broad applications in drug discovery, protein engineering, and force field development.

Dedication

To my beloved wife, Mahsan, whose love and support made this journey possible. And to my family, for their endless encouragement and belief in me.

Acknowledgements

I would like to express my deepest gratitude to all those who supported me throughout the journey of completing this thesis.

First and foremost, I thank my wonderful wife, Mahsan, for her unwavering love, patience, and encouragement. Her support has been the foundation of my strength during the most challenging moments of this work. I am also deeply grateful to my family for their unconditional love and belief in me from the very beginning.

I would like to extend my sincere thanks to my supervisor, Dr. Stijn De Baerdemacker, for his invaluable guidance, mentorship, and continuous support. His insights and dedication have been instrumental in shaping this research.

I am also thankful to the QuNB group for the insightful discussions and collaborative spirit that enriched my academic experience. Special thanks to the Department of Chemistry and the University of New Brunswick for providing a supportive and stimulating environment for research.

To everyone who contributed to this journey in ways big or small, thank you.

Table of Contents

Abstract ii

Dedication iv

Acknowledgments v

Table of Contents vi

List of Tables xi

List of Figures xii

Abbreviations xiv

1 Introduction 1

1.1 Significance of Biomolecular Structures 1

1.1.1 Proteins: The Building Blocks of Life 1

1.1.2 Experimental Techniques for Protein Structure Determination 5

1.1.2.1 X-ray Crystallography 5

1.1.2.2 Nuclear Magnetic Resonance Spectroscopy 6

1.1.2.3 Cryo-Electron Microscopy 6

1.1.3 Challenges in Protein Folding 6

1.2 Theoretical Framework for Biomolecular Modeling 9

1.2.1 Classical Mechanics 9

1.2.2 Semi-Empirical Methods 11
1.2.3 Quantum Mechanics 12
1.2.3.1 Born-Oppenheimer Approximation 14
1.2.3.2 The Many-Body Problem 15
1.2.3.3 Electronic Structure Methods 17
1.2.3.4 Hartree-Fock Theory 17
1.2.3.5 The Linear Combination of Atomic Orbitals 18
1.2.3.6 Fock Operator 18
1.2.3.7 Variational Principle 19
1.2.3.8 Basis Functions 20
1.2.3.9 Second Quantization 22
1.2.3.10 Wick's Theorem 24
1.2.3.11 Density Functional Theory 25
1.2.3.12 Quantum Theory of Atoms in Molecules 27
1.2.3.13 Transferable Atom Equivalent 28
1.2.3.14 Fragment Molecular Orbital 29
1.2.3.15 Kernel Energy Method 30
1.2.3.16 Energy Decomposition Analysis 31
1.2.3.17 Natural Bond Orbital Analysis 31
1.2.3.18 Size Extensiveness 32
1.2.3.19 Post-Hartree-Fock Methods 32
1.2.3.20 Configuration Interaction 32
1.2.3.21 Møller-Plesset Perturbation Theory 33
1.2.3.22 Multi Reference Methods 33
1.2.3.23 Coupled Cluster Theory 34
1.2.3.24 Density Matrix Renormalization Group 35
1.2.3.25 Locality in Electronic Structure 40

1.2.3.26 Density Matrix Formalism 41
1.2.3.27 Reduced Density Matrices 41
1.2.3.28 Quantum Entanglement 42
1.2.3.29 One-Orbital Reduced Density Matrix 42
1.2.3.30 Two-Orbital Reduced Density Matrix 45
1.2.3.31 Mutual Information 47
1.2.3.32 Atomic Mutual Information 47
1.2.3.33 Fragment Mutual Information 49
1.2.3.34 Localized Molecular Orbitals 50
1.2.3.35 Canonical Orthogonalization 52
1.3 Outline of the Thesis 53
Bibliography 55

2 Correlation in Dipeptide Systems from Quantum Information Theory Perspectives 67
2.1 Introduction 68
2.2 Methods 75
2.2.1 Mutual information 75
2.2.2 Dipeptides preparation and computational details 82
2.3 Results and discussions 83
2.4 Conclusion 97
2.5 Acknowledgements 99
2.6 Appendix 99
2.6.1 Mutual Information from Slater determinant wavefunctions . . 99
2.6.2 DMRG Calculations 101
Bibliography 105

3 From dipeptide systems to polypeptides: evolution of mutual information. 119

3.1 Introduction . . . 120

3.2 Methodology . . . 123

3.2.1 Calculation of Mutual Information . . . 123

3.2.2 Computational Workflow of the dipeptides . . . 125

3.2.3 Computational Workflow of the polypeptide . . . 126

3.3 Results and discussions . . . 128

3.4 Conclusion . . . 142

3.5 Acknowledgements . . . 144

Bibliography . . . 144

4 Mapping Electronic Correlations in Proteins via Cut-Wise Reconstruction: A Case Study on Insulin 153

4.1 Introduction . . . 154

4.2 Methodology . . . 156

4.2.1 Mutual Information from Reduced Density Matrices . . . 156

4.2.2 From Orbital to Atomic and Residue-Level Correlations . . . 157

4.2.3 System Preparation and Fragmentation . . . 158

4.3 Results and discussions . . . 159

4.4 Conclusion . . . 166

4.5 Acknowledgements . . . 167

Bibliography . . . 167

5 Discussion 172

5.1 From Orbitals to Atoms . . . 172

5.2 From Atoms to Residues . . . 177

5.3 Toward Larger Scales . . . 180

5.4 Future works 181
Bibliography 183

Vita

List of Tables

1.1 Complete set of single-orbital transition operators 43

1.2 Mapping of the diagonal entries of the ρ_i. 44

1.3 Sixteen single-orbital operators. 45

1.4 2-RDM in the occupation number basis. 46

3.1 Hydrogen Bonds and Correlation in Metastable Structures 139

4.1 List of hydrogen bonds in insulin 165

List of Figures

1.1 Amino acid structure and codes 2
1.2 Peptide bond formation 3
1.3 Protein structure levels 4
1.4 Basis functions in quantum chemistry 22
1.5 Matrix Product State (MPS) 38
1.6 Canonical form of a MPS 39

2.1 1- and 2-orbitals RDMs 78
2.2 Atomic Mutual Information (AMI) 80
2.3 Circular vs. Spatial Atom-Wise Networking Plot (SPAWN) 84
2.4 SPAWN plots for 12 dipeptides 85
2.5 Glycine-Glycine dipeptide SPAWN plot 86
2.6 AMI analysis in stretched bonds 88
2.7 Pairwise AMI for 12 dipeptides 89
2.8 AMI vs. distance in Ala tetra-peptide 91
2.9 SPAWN plots of Gly-Gly at different levels of theory 92
2.10 AMI comparisons of Gly-Gly at different levels of theory 93
2.11 AMI evolution in Cyanide during hydrogen dissociation 95
2.12 SPAWN plots for cyanide and formaldehyde 96
2.13 AMI vs. basis set size in CysGly 97
2.14 AMI plot for Ala-Gly obtained from DMRG 101
2.15 DMRG ($D = 1024, 2048$) derived MI for GlyGly and AlaGly 102

2.16 MI and energy vs. grid resolution 103
2.17 AMI plots for different localized orbital bases 104

3.1 Atomization energy vs. AMI 129
3.2 FMI and EDA across dipeptides 131
3.3 FMI vs. bonding energy correlation 131
3.4 AMI vs. Lennard-Jones energy in Lys-Lys and Ser-Gly 133
3.5 AMI-LJ correlation in CHARMM and OPLS-AA 134
3.6 Energy and FMI Evolution During Neh2 Folding 135
3.7 Energy and FMI Evolution During Neh2 Folding 137
3.8 Resolution Difference Between FMI and AMI 138
3.9 Neh2 Conformations and Corresponding SPAWN Plots 140
3.10 Classical vs. DFT FMI SPAWN Plot 142

4.1 Insulin structure and fragments used in QM calculations 160
4.2 AMI SPAWN plots from cut-wise vs. full DFT in insulin 161
4.3 FMI SPAWN plots from cut-wise vs. full DFT in insulin 162
4.4 FMI heatmaps of cut-wise vs. full DFT in insulin 164
4.5 Scatter plot cut-wise vs. full DFT in insulin 166

List of Symbols, Nomenclature or Abbreviations

AIM	Atom in a Molecule
AM1	Austin Model 1
AMI	Atomic Mutual Information
AO	Atomic Orbital
BPC	Bond Critical Points
CASSCF	Complete Active Space Self-Consistent Field
CASPT2	Complete Active Space Second-Order Perturbation Theory
cc-pVXZ	Correlation-Consistent Polarized Valence X-Zeta (X = D, T, Q, 5, ...)
CC	Coupled Cluster
CGTO	Contracted Gaussian-Type Orbital
CI	Configuration Interaction
CISD	CI Singles and Doubles
DFT	Density Functional Theory
DLPNO	Domain-Based Local Pair Natural Orbital
DMRG	Density Matrix Renormalization Group
EDA	Energy Decomposition Analysis
ER	Edmiston–Ruedenberg
FB	Foster–Boys

FCI	Full Configuration Interaction
FMI	Fragment-wise Mutual Information
GGA	Generalized Gradient Approximation
GTO	Gaussian-Type Orbital
HF	Hartree-Fock
LDA	Local Density Approximation
LJ	Lennard-Jones
LMO	Localized Molecular Orbital
LMOs	Localized Molecular Orbitals
MD	Molecular Dynamics
MI	Mutual Information
MO	Molecular Orbital
MPn	Møller-Plesset Perturbation Theory (n-th order)
MPS	Matrix Product State
NBO	Natural Bond Orbital
NMR	Nuclear Magnetic Resonance
PDB	Protein Data Bank
PM	Pipek–Mezey
PM3	Parametric Method 3
PNO	Pair Natural Orbital
QTAIM	Quantum Theory of Atoms in Molecules
QM	Quantum Mechanics
QIT	Quantum Information Theory
RDM	Reduced Density Matrix
RHF	Restricted Hartree-Fock
SPAWN	Spatial Pairwise Atomic Wiring Network
STO	Slater-Type Orbital

Chapter 1

Introduction

1.1 Significance of Biomolecular Structures

Biological systems rely on a wide variety of macromolecules to perform essential functions. Among these, proteins play a particularly critical role as the primary molecular machinery of life. Understanding the structure and function of proteins is fundamental to uncovering the mechanisms that govern biological processes and addressing diseases caused by their dysfunction.

1.1.1 Proteins: The Building Blocks of Life

Proteins are composed of amino acids, which are organic molecules serving as the building blocks of these essential macromolecules. Each amino acid consists of a central carbon atom (called the $\alpha-$carbon) bonded to four distinct groups: a hydrogen atom, a carboxyl group ($-$COOH), an amino group ($-NH_2$), and a variable side chain (R-group), refer to Figure 1.2. The unique properties of the R-group determine the chemical nature of each amino acid, such as being hydrophobic, hydrophilic, acidic, or basic. Amino acids exhibit chirality, meaning they exist in two mirror-image forms known as enantiomers: the L- and D-forms. In biological systems, proteins are exclu-

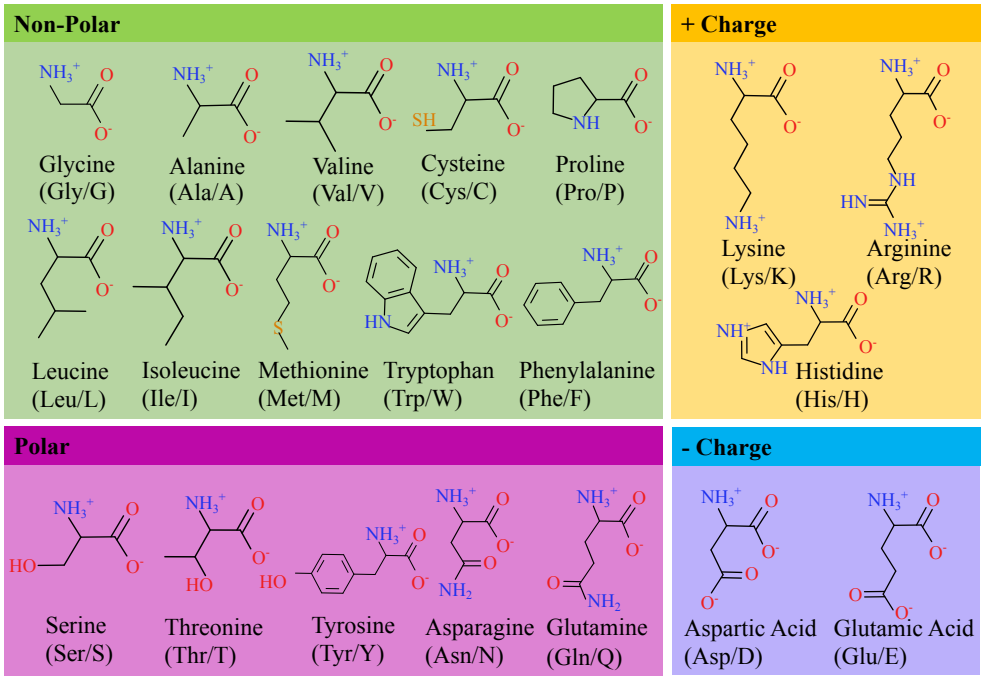


Figure 1.1: Structures of amino acids showing their common backbone, variable side chains, and corresponding three-letter and one-letter codes.

sively composed of L-amino acids [1]. There are 20 naturally occurring amino acids, and their specific sequence in a polypeptide chain dictates the protein's structure and function, Figure 1.1. Amino acids are linked together through peptide bonds, where the carboxyl group of one amino acid reacts with the amino group of another, releasing a molecule of water in a condensation reaction, Figure 1.2. The resulting bond links the amino acids into a linear chain called a polypeptide. This process is catalyzed by ribosomes in living cells during protein synthesis [2]. The sequence of amino acids in the polypeptide determines its primary structure and ultimately influences the folding and function of the protein. The folded structure determines the specific function of the protein, enabling it to act as an enzyme, structural components, signaling molecules, transporters, or antibodies within biological systems. The structure of a protein can be organized into hierarchical levels [3]:

- **Primary Structure**: The linear sequence of amino acids in a polypeptide chain, linked by peptide bonds. This sequence determines the subsequent fold-

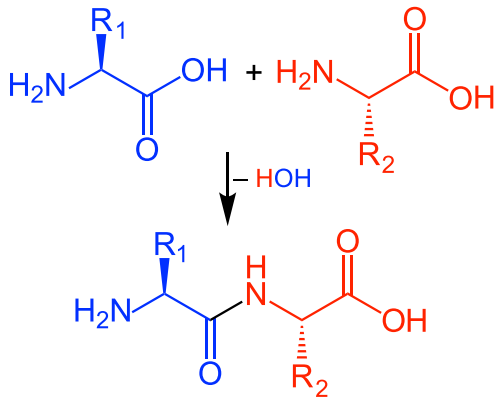


Figure 1.2: peptide bond formation between two amino acids, illustrating the condensation reaction that results in a polypeptide chain

ing and overall structure of the protein, see Figure 1.3.

- **Secondary Structure**: Localized folding patterns stabilized by hydrogen bonds between the backbone atoms of the polypeptide chain. Common secondary structures include α-helices and β-sheets.

- **Tertiary Structure**: The overall three-dimensional arrangement of a single polypeptide chain is determined by interactions between side chains and the surrounding environment.

- **Quaternary Structure (if applicable)**: The assembly of multiple polypeptide chains into a functional protein complex stabilized by non-covalent interactions or disulfide bonds.

These levels of organization are stabilized by noncovalent interactions, including hydrogen bonds, van der Waals forces, and hydrophobic interactions, see Figure 1.3. The precise folding of a protein enables it to form functional domains, active sites, and binding pockets critical for its role in biological systems. Any disruption to this structure, whether due to environmental changes or genetic mutations, can affect protein function and may result in a process known as denaturation. During denaturation, the protein unfolds or loses its organized structure. Unlike degradation,

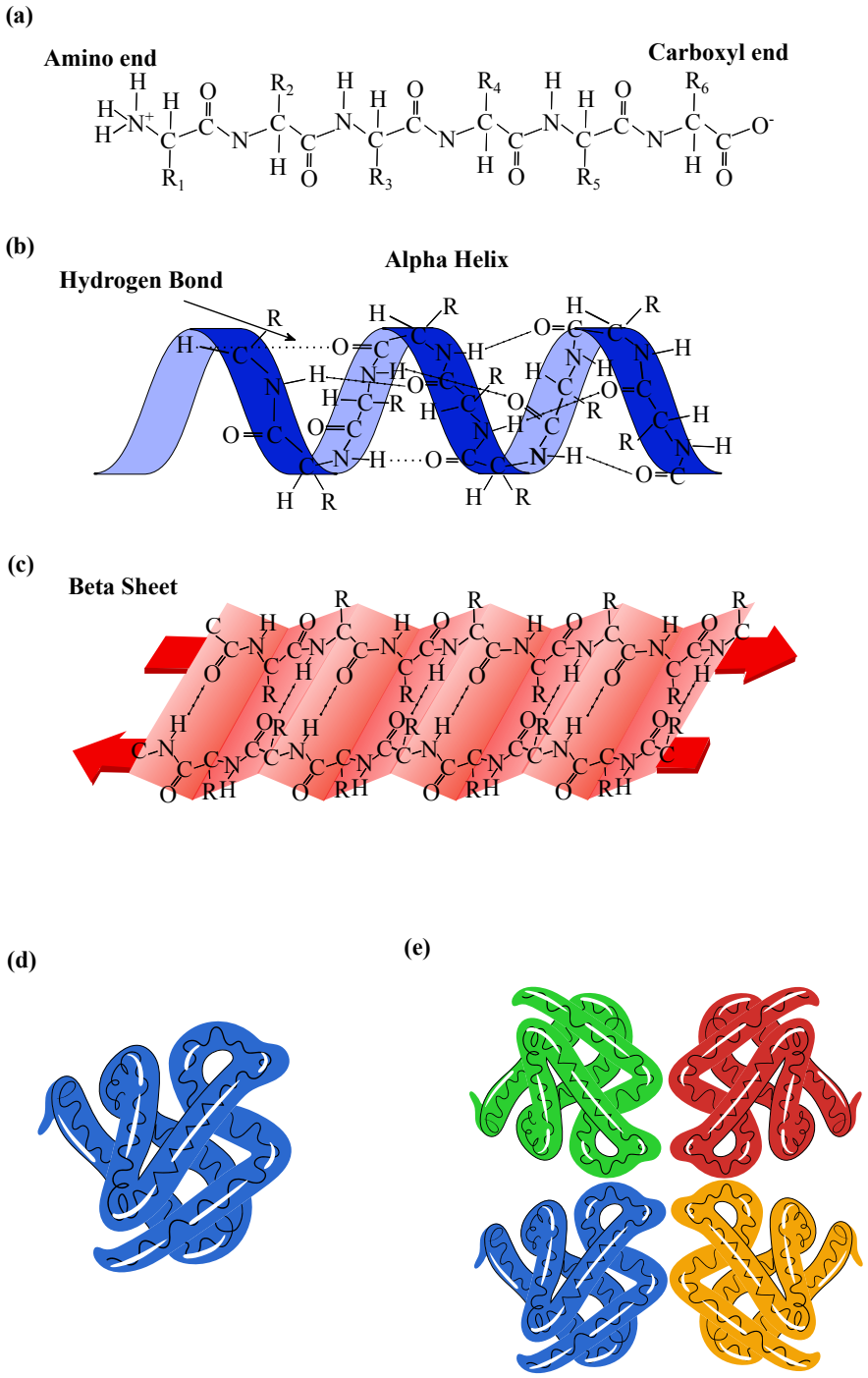


Figure 1.3: Cartoon representation of the hierarchical levels of protein structure, illustrating (a) primary (amino acid sequence), secondary (b) α-helices and (c) β-sheets, (d) tertiary (three-dimensional folding), and (e) quaternary (assembly of multiple polypeptide chains) structures.

which breaks the peptide bonds, denaturation affects non-covalent interactions and the folding of the polypeptide chain. While denaturation is often irreversible, in some cases, proteins can refold into their native conformation under favorable conditions [3].

1.1.2 Experimental Techniques for Protein Structure Determination

Understanding protein structures at an atomic level is essential for deciphering their functions and interactions within biological systems. Several experimental techniques have been developed to determine protein structures with high precision. The three most widely used methods are X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM).

Each of these methods has unique strengths and limitations, and researchers often use complementary approaches to gain a comprehensive understanding of protein structures and dynamics. These experimental techniques, combined with computational modeling, continue to expand our knowledge of biomolecular structures and their roles in biological processes.

1.1.2.1 X-ray Crystallography

X-ray crystallography has historically been the most common method for determining protein structures at atomic resolution. This technique involves crystallizing the protein of interest and exposing it to a beam of X-rays. The X-rays scatter upon interacting with the electron clouds of the atoms in the crystal, which produces a diffraction pattern that is recorded and analyzed to reconstruct the three-dimensional structure of the protein [4]. The success of X-ray crystallography depends on the ability to obtain high-quality crystals, a challenge for certain proteins, particularly membrane proteins and intrinsically disordered proteins. Despite this limitation, X-

ray crystallography has provided the majority of high-resolution protein structures available in the Protein Data Bank (PDB).

1.1.2.2 Nuclear Magnetic Resonance Spectroscopy

NMR spectroscopy is a powerful alternative to X-ray crystallography for studying proteins in solution, allowing structural characterization without the need for crystallization. NMR relies on the magnetic properties of atomic nuclei, particularly hydrogen (^{1}H), carbon (^{13}C), and nitrogen (^{15}N), to probe the local chemical environment of atoms within the protein [5]. By analyzing the interactions between nuclear spins, NMR spectroscopy provides information on protein structure, dynamics, and interactions in physiological conditions. However, NMR is limited to relatively small proteins (typically less than 40 kDa) due to the complexity of spectral analysis.

1.1.2.3 Cryo-Electron Microscopy

Cryo-electron microscopy has revolutionized structural biology by enabling the determination of protein structures at near-atomic resolution without the need for crystallization. In cryo-EM, proteins are rapidly frozen in a thin layer of vitreous ice and imaged using an electron microscope. Computational methods then reconstruct a three-dimensional model from thousands of individual images [6]. Cryo-EM is particularly advantageous for studying large protein complexes and membrane proteins, which are challenging to crystallize or analyze using NMR. The technique has seen remarkable advances in recent years, making it a crucial tool in modern structural biology.

1.1.3 Challenges in Protein Folding

The folding of proteins into their functional conformations is a complex process. After synthesis, the linear polypeptide chain must navigate a vast conformational landscape

to achieve its native three-dimensional structure. This process is governed by the amino acid sequence, as described by Anfinsen's dogma [7]. Anfinsen demonstrated that the native structure, a state of lowest free energy, of a protein is determined solely by its primary sequence of amino acids. This principle underscores the deterministic relationship between sequence and structure [8]. The complexity of protein folding arises from the immense number of possible conformations a polypeptide chain can adopt. Levinthal's paradox illustrates this challenge [9]: if a protein sampled every possible conformation randomly, the time required to fold would exceed the age of the universe. Yet, proteins typically fold within milliseconds to seconds. This apparent paradox is resolved by the concept of a folding funnel, which describes the energy landscape of folding. In this model, proteins progress through a series of intermediate states, guided by a decreasing free energy gradient, until they reach their native conformation. Understanding protein folding is essential not only for elucidating fundamental biological processes but also for addressing diseases caused by misfolding, where the failure to attain the correct structure results in loss of function or toxic aggregation. In Alzheimer's disease, misfolding of the amyloid-β protein leads to the formation of plaques in the brain, which are associated with cognitive decline [10]. Similarly, in Parkinson's disease, aggregation of α-synuclein forms Lewy bodies, which disrupt neuronal function [11]. The cellular stress, inflammation, and apoptosis triggered by misfolded proteins exacerbate these conditions, emphasizing the importance of understanding the folding process [12]. To predict the three-dimensional structure of proteins and better understand their folding mechanisms, researchers have developed a variety of computational and theoretical models. One such model is the Go model, a coarse-grained approach that assumes the native state of a protein is its most thermodynamically stable conformation [13]. By focusing on native contacts defined in experimentally determined structures, the Go model provides valuable insights into folding pathways and kinetics. However, it relies on prior knowledge of

the native structure, limiting its applicability to unknown systems [14].

Molecular dynamics (MD) simulations using coarse-grained force fields, such as MARTINI [15] or AWSEM [16], have also been employed to study protein folding. These models simplify the representation of proteins by grouping atoms into larger beads, reducing computational cost while maintaining key structural features. Coarse-grained MD simulations enable the study of larger systems and longer timescales, capturing aspects such as secondary structure formation and global dynamics that are challenging to model with all-atom approaches.

A groundbreaking advancement in protein structure prediction has come from AlphaFold, developed by DeepMind [17, 18]. AlphaFold utilizes deep learning techniques to predict protein structures with remarkable accuracy by analyzing large-scale sequence and structural data. Its neural networks integrate information from multiple sequence alignments and structural databases, achieving atomic-level precision in predicting protein conformations. In 2021, AlphaFold successfully predicted the structures of nearly all human proteins, a milestone that has revolutionized the field of structural biology [19]. This achievement has profound implications for biomedical research, accelerating drug discovery, enabling the study of previously uncharacterized proteins, and providing insights into diseases caused by structural defects. The significance of AlphaFold's contribution was recognized with the Nobel Prize in 2024 [20].

These approaches, ranging from physics-based models like the Go model and coarse-grained simulations to data-driven methods like AlphaFold, demonstrate the diverse strategies employed to understand protein folding. Each method contributes uniquely to unraveling the complexities of the folding process, and their integration with experimental data provides a comprehensive framework for investigating the relationship between protein structure, function, and disease.

1.2 Theoretical Framework for Biomolecular Modeling

Understanding biomolecular systems requires a combination of theoretical approaches, each suited to different scales of complexity and computational cost. These approaches can be broadly categorized into three levels of theory: classical mechanics, semi-empirical methods, and quantum mechanics. This section introduces these levels, starting with classical mechanics and progressing to the advanced precision of quantum mechanics.

1.2.1 Classical Mechanics

Classical mechanics is the fundamental framework that describes the motion of objects using Newton's laws. It provides a deterministic approach where the trajectory of a system can be predicted if its initial conditions and the forces acting on it are known. In this framework, atomic and molecular systems are treated as collections of classical particles, where interactions between them arise from potential energy functions. These energy functions dictate how forces influence the motion of the particles in a chemical system.

One important application of classical mechanics is MD, a computational technique used to simulate the motion of atoms and molecules over time. MD solves Newton's equations of motion.

$$\vec{F_i} = m_i \frac{d^2 \vec{r_i}}{dt^2} \tag{1.1}$$

where $\vec{F_i}$ is the force on atom i, m_i is its mass, and $\vec{r_i}$ is its position. By tracking the positions and velocities of particles, MD allows us to study molecular behavior, conformational changes, and dynamic interactions in biological and chemical systems.

This approach is particularly useful in fields like drug discovery, protein folding, and biophysics, where understanding atomic-level processes is essential [21].

The accuracy and efficiency of MD simulations depend on the force fields, which define the potential energy of a system based on a set of mathematical functions and parameters. They approximate the interactions between atoms and molecules by considering bonded interactions, such as covalent bonds modeled with harmonic potentials, and non-bonded interactions, including van der Waals forces and electrostatic interactions [22, 23]. The total energy E_{total} of a molecular system is expressed as:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{non-bonded}} \tag{1.2}$$

where E_{bonded} includes bond stretching, angle bending, and torsional terms, while $E_{\text{non-bonded}}$ accounts for van der Waals and electrostatic interactions.

for the bonded term we have:

$$E_{\text{bonded}} = \sum_{\text{bonds}} k_b(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi[1 + \cos(n\phi - \phi_0)] \tag{1.3}$$

where k_b, k_θ, k_ϕ are force constants for bonds, angles, and torsions, respectively. These force constants are typically derived from quantum mechanical calculations, experimental data (such as vibrational spectra), or empirical fitting to reproduce observed molecular properties. r, θ, ϕ represent bond lengths, bond angles, and dihedral angles, while r_0, θ_0, ϕ_0 are their corresponding equilibrium values, obtained from high-quality quantum chemistry calculations or experimental crystallographic data.

The non-bonded interactions consist of van der Waals forces and electrostatic interactions. The van der Waals forces are often modeled using the Lennard-Jones (LJ) potential, which describes the interaction between two non-bonded atoms or

molecules. The LJ potential is expressed as:

$$E_{\mathrm{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{1.4}$$

Here ϵ is the depth of the potential well, representing the strength of the interaction. σ is the distance at which the potential is zero, related to the size of the particles, and r is the distance between the two interacting particles.

The LJ potential balances two competing forces [24]. The $\frac{-1}{r^6}$ term represents the attractive van der Waals interaction, also known as London dispersion forces. The $\frac{1}{r^{12}}$ term represents the repulsive interaction and serves as an empirical approximation of the Pauli exclusion principle.

LJ has proven to be an excellent approximation for modeling nonbonded interactions in many systems, particularly in molecular simulations. Including the electrostatic term, the full expression for nonbonded interactions is:

$$E_{\mathrm{nonbonded}} = \sum_{i<j} \left[4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right] \tag{1.5}$$

Where q_i, q_j are partial charges of atoms i and j and ϵ_0 is permittivity of free space. Force fields such as CHARMM [25], and OPLS-AA [26] rely on the LJ potential for van der Waals interactions and Coulomb's law for electrostatics. These parameters are tuned to reproduce experimental observations and quantum mechanical calculations, making force fields a versatile tool for molecular modeling.

1.2.2 Semi-Empirical Methods

Semi-empirical methods combine principles of quantum mechanics with parameterization from experimental or high-level theoretical data. These methods simplify quantum mechanical calculations by neglecting certain integrals and approximating

others, reducing the computational cost. Some commonly used semi-empirical methods include the Austin Model 1 (AM1) [27] and Parametric Method 3 (PM3) [28]. Semi-empirical approaches are particularly useful for studying systems where electronic interactions play a significant role but full quantum mechanical treatment is computationally prohibitive. They are often used for exploring reaction mechanisms, charge transfer processes, and modeling active sites in an enzyme. Despite their advantages, semi-empirical methods are not without limitations. Their reliance on parameterization restricts their applicability to systems similar to those used during the parameterization process [29].

1.2.3 Quantum Mechanics

Quantum mechanics provides the most detailed framework for studying molecular systems by explicitly considering the electronic structure. It is governed by the time-independent Schrödinger equation [30, 31]:

$$\hat{H}\psi = E\psi \tag{1.6}$$

where $\hat{H}$ is the Hamiltonian operator, ψ is the wavefunction of the system, and E represents the total energy of the system. The Hamiltonian operator, $\hat{H}$, encapsulates all energy components of the molecular system and is expressed as:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{nn} + \hat{V}_{ne} + \hat{V}_{ee} \tag{1.7}$$

Each term represents a distinct physical contribution to the total energy:
$\hat{T}_n$ term accounts for the kinetic energy of the nuclei, which are treated as massive particles compared to electrons. It is expressed mathematically as:

$$\hat{T}_n = -\sum_I \frac{\hbar^2}{2M_I}\nabla_I^2 \tag{1.8}$$

where M_I is mass of the I-th nucleus. ∇_I^2 is Laplacian operator acting on the nuclear coordinates, $\hbar$ is reduced Planck's constant.

$\hat{T}_e$ term accounts for the kinetic energy of all electrons in the system. It is expressed as:

$$\hat{T}_e = -\sum_i \frac{\hbar^2}{2m_e}\nabla_i^2 \tag{1.9}$$

where m_e is the mass of the electron. ∇_i^2 is Laplacian operator acting on the position of the i-th electron.

$\hat{V}_{nn}$ term describes the repulsive Coulombic interactions between positively charged nuclei in the system. It is expressed as:

$$\hat{V}_{nn} = \sum_{I<J} \frac{Z_I Z_J e^2}{R_{IJ}} \tag{1.10}$$

where Z_I and Z_J are atomic numbers of nuclei I and J, e is elementary charge. R_{IJ} is the distance between nuclei I and J.

$\hat{V}_{ne}$ term represents the attractive Coulombic interactions between the positively charged nuclei and negatively charged electrons. It is given by:

$$\hat{V}_{ne} = -\sum_i \sum_I \frac{Z_I e^2}{r_{iI}} \tag{1.11}$$

where r_{iI} is the distance between the i-th electron and the I-th nucleus.

$\hat{V}_{ee}$ term accounts for the repulsive interactions between pairs of electrons due to their negative charge. It is expressed as:

$$\hat{V}_{ee} = \sum_{i<j} \frac{e^2}{r_{ij}} \tag{1.12}$$

where r_{ij} is distance between electrons i and j.

1.2.3.1 Born-Oppenheimer Approximation

The Born-Oppenheimer approximation simplifies the molecular Schrödinger equation by decoupling the motions of nuclei and electrons. Since nuclei are significantly heavier and move much more slowly than electrons, their dynamics can be treated separately [32]. This approximation leads to the factorization of the total molecular wavefunction into electronic and nuclear components:

$$\psi(\mathbf{R}, \mathbf{r}) \approx \psi_e(\mathbf{r}; \mathbf{R})\chi_n(\mathbf{R}) \tag{1.13}$$

where $\psi_e(\mathbf{r}; \mathbf{R})$ is the electronic wavefunction, dependent on the positions of electrons ($\mathbf{r}$) and parametrically on the nuclear positions ($\mathbf{R}$).$\chi_n(\mathbf{R})$ is the nuclear wavefunction, which describes the motion of the nuclei.

The Born-Oppenheimer approximation allows the electronic Schrödinger equation to be solved for fixed nuclear positions, significantly reducing the computational complexity of molecular systems. The resulting electronic energy $E_{\text{el}}(\mathbf{R})$ defines the potential energy surface (PES), which in turn governs the nuclear motion.

After applying the Born-Oppenheimer approximation, the total molecular Hamiltonian can be separated into two main components: the electronic Hamiltonian and the nuclear Hamiltonian. This separation simplifies the description of molecular systems by treating electronic and nuclear motions independently to a good approximation:

$$\hat{H} = \hat{T}_{\text{n}} + E_{\text{el}}(\mathbf{R}) \tag{1.14}$$

where $\hat{T}_{\text{n}}$ is the nuclear Hamiltonian, which governs the motion of the nuclei on the potential energy surface. $E_{\text{el}}(\mathbf{R})$ is the electronic energy, obtained by solving the electronic Hamiltonian for fixed nuclear positions.

It is important to note that the term $E_{\text{el}}(\mathbf{R})$ in the total Hamiltonian represents the

eigenvalue of the electronic Hamiltonian $\hat{H}_{\text{el}}$, not the Hamiltonian operator itself. The electronic Schrödinger equation:

$$\hat{H}_{\text{el}}\psi_e(\mathbf{r};\mathbf{R}) = E_{\text{el}}(\mathbf{R})\psi_e(\mathbf{r};\mathbf{R}), \tag{1.15}$$

Where $\psi_e(\mathbf{r};\mathbf{R})$ encodes the spatial distribution of electrons for fixed nuclear positions is solved for each fixed nuclear configuration $\mathbf{R}$, and the resulting eigenvalue $E_{\text{el}}(\mathbf{R})$ acts as a scalar potential energy term. The electronic Hamiltonian is given by:

$$\hat{H}_{\text{el}} = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} \tag{1.16}$$

From this point onward, we will focus on the electronic Hamiltonian $\hat{H}_{\text{el}}$ and its role in describing the electronic structure of molecular systems. The simplifications provided by the Born-Oppenheimer approximation allow us to isolate the electronic contributions, which form the foundation of modern quantum chemistry methods.

1.2.3.2 The Many-Body Problem

The many-body problem in quantum mechanics refers to the challenge of solving the Schrödinger equation for systems with multiple interacting particles, such as electrons in atoms and molecules. The time-independent Schrödinger equation is expressed as:

$$\hat{H}\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N), \tag{1.17}$$

where $\hat{H}$ is the Hamiltonian operator, Ψ is the wavefunction of the system, E is the total energy, and N is the number of particles.

The Hamiltonian for a many-body system can be written as:

$$\hat{H} = \sum_{i=1}^{N} \hat{T}_i + \sum_{i<j}^{N} \hat{V}_{ij} + E_o, \tag{1.18}$$

where $\hat{T}_i$ represents the kinetic energy of the i-th particle and $\hat{V}_{ij}$ describes the interactions between particles i and j, E_o is a constant, e.g. V_{nn} for $\hat{H}_{\text{el}}$.

One of the primary challenges in the many-body problem is the exponential growth of degrees of freedom with the number of particles. The wavefunction depends on the spatial coordinates of all N particles, making direct computation a combinatorial problem. This complexity arises because the motion of one particle is influenced by the presence of others, particularly due to mutual Coulomb repulsion. This interdependence of particle motions is referred to as electron correlation, a critical aspect of the many-body problem.

Electron correlation can be broadly classified into two types: dynamic and static. Dynamic correlation accounts for the short-range interactions between electrons as they avoid each other to minimize repulsion, capturing instantaneous fluctuations in their positions. In contrast, static correlation becomes significant in systems where multiple electronic configurations are nearly degenerate, such as during bond-breaking processes or in transition states, and is essential for describing the correct ground state.

Despite significant advancements, the many-body problem and its associated correlation effects remain fundamental challenges in theoretical and computational chemistry and physics. In quantum chemistry, they are indispensable for understanding molecular interactions, electronic structures, and reaction mechanisms. In condensed matter physics, they provide insights into phenomena such as (super)conductivity, magnetism, and others. Efforts to address these challenges have led to the development of numerous computational and theoretical methods, which will be explored in detail in the section on electronic structure methods.

1.2.3.3 Electronic Structure Methods

Electronic structure methods are the cornerstone of computational chemistry, developed to approximate solutions to the Schrödinger equation 1.15 for molecular systems. These methods provide insights into the behavior of electrons within molecules, revealing critical information about molecular properties such as total energy, molecular orbitals (MOs), bond strengths, dipole moments, and chemical reactivity. By describing how electrons interact with nuclei and with each other, electronic structure methods bridge the gap between quantum theory and practical applications in chemistry, biology, and materials science.

Due to the inherently complex nature of the many-body problem, solving the Schrödinger equation accurately is only feasible for very small systems, typically containing fewer than ten particles. This limitation arises from the exponential growth in the number of electronic interactions as the system size increases. For larger and more complex systems, approximations and numerical techniques are essential. These methods aim to balance computational efficiency with accuracy, enabling researchers to study realistic molecular systems.

1.2.3.4 Hartree-Fock Theory

Hartree-Fock (HF) theory is a foundational method in quantum chemistry, developed to approximate solutions to the electronic Schrödinger equation for many-electron systems. HF theory forms the foundation for many advanced quantum chemical methods and is essential for understanding the behavior of electrons in molecules. HF theory addresses the many-electron problem by assuming that each electron moves independently in an average field created by all other electrons. This independent particle approximation reduces the complexity of the problem significantly.

The foundation of HF theory lies in the variational principle, which ensures that the approximate wavefunction minimizes the total electronic energy. In HF theory, the

wavefunction is represented as a single Slater determinant:

$$\Psi(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \ldots, \mathbf{r}_N, s_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_1(\mathbf{r}_1, s_1) & \psi_2(\mathbf{r}_1, s_1) & \ldots & \psi_N(\mathbf{r}_1, s_1) \\ \psi_1(\mathbf{r}_2, s_2) & \psi_2(\mathbf{r}_2, s_2) & \ldots & \psi_N(\mathbf{r}_2, s_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{r}_N, s_N) & \psi_2(\mathbf{r}_N, s_N) & \ldots & \psi_N(\mathbf{r}_N, s_N) \end{vmatrix} \tag{1.19}$$

where $\psi_i(\mathbf{r}, s)$ are the spin-orbitals, each containing a spatial molecular orbital (MO) component and a spin function. Here, r represents the spatial coordinates of an electron, while s denotes its spin variable. The Slater determinant ensures compliance with the Pauli exclusion principle and accounts for exchange interactions.

1.2.3.5 The Linear Combination of Atomic Orbitals

Since molecular orbitals are not known a priori, they are constructed using the Linear Combination of Atomic Orbitals (LCAO) approximation:

$$\psi_i(\mathbf{r}, s) = \sum_{\mu} c_{\mu i} \phi_\mu(\mathbf{r}) \chi(s), \tag{1.20}$$

where $\psi_i(\mathbf{r}, s)$ is the i-th spin-orbital, $\phi_\mu(\mathbf{r})$ are atomic orbitals (AOs), which include a spatial component and a spin component, $\chi(s)$, which can be either $\alpha(s)$ (spin-up) or $\beta(s)$ (spin-down), $c_{\mu i}$ are the molecular orbital coefficients. The coefficients $c_{\mu i}$ are determined by solving the HF equations.

1.2.3.6 Fock Operator

Minimizing the total electronic energy leads to the HF equations:

$$\hat{F}\psi_i = \varepsilon_i \psi_i, \tag{1.21}$$

where $\hat{F}$ is the Fock operator, defined as:

$$\hat{F} = \hat{h} + \hat{J} - \hat{K}, \tag{1.22}$$

Here $\hat{h}$ is the core Hamiltonian includes the one-electron terms, representing the kinetic energy of the electron (Eq. 1.9) and its attraction to the nuclei (Eq. 1.11). The $\hat{J}$ is the Coulomb operator that accounts for the average coulomb repulsion between an electron and the total electron density:

$$\hat{J}_j\psi_i(\mathbf{r}_1, s_1) = \left[\int \frac{|\psi_j(\mathbf{r}_2, s_2)|^2}{r_{12}} d\mathbf{r}_2 ds_2\right] \psi_i(\mathbf{r}_1, s_1), \tag{1.23}$$

where r_{12} is the distance between electrons 1 and 2. The exchange operator $\hat{K}$ arises due to the antisymmetry of the wavefunction, ensuring compliance with the Pauli exclusion principle:

$$\hat{K}_j\psi_i(\mathbf{r}_1, s_1) = \left[\int \frac{\psi_j^*(\mathbf{r}_2, s_2)\psi_i(\mathbf{r}_2, s_2)}{r_{12}} d\mathbf{r}_2 ds_2\right] \psi_j(\mathbf{r}_1, s_1). \tag{1.24}$$

1.2.3.7 Variational Principle

The HF method leverages the variational principle to approximate the electronic wavefunction. According to the variational principle, the ground-state energy E_0 of a quantum system satisfies:

$$E_0 \leq \frac{\langle\Psi|\hat{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{1.25}$$

The goal is to find the wavefunction Ψ that minimizes the total energy:

$$E_{\text{HF}} = \min_{\psi_1,\psi_2,\ldots,\psi_N} \left(\sum_{i=1}^{N}\langle\psi_i|\hat{h}|\psi_i\rangle + \frac{1}{2}\sum_{i,j=1}^{N}\left(\langle\psi_i\psi_j|\hat{J}|\psi_i\psi_j\rangle - \langle\psi_i\psi_j|\hat{K}|\psi_i\psi_j\rangle\right)\right) + V_{nn}. \tag{1.26}$$

1.2.3.8 Basis Functions

Atomic orbitals were initially approximated using Slater-Type Orbitals (STOs) [33], which resemble the solutions of the hydrogen-like Schrödinger equation

$$\phi_{STO}(r) = e^{-\alpha r}, \tag{1.27}$$

where α is a constant related to the effective charge of the nucleus and r is the radial distance from the nucleus. STOs correctly capture the nuclear cusp behavior of atomic wavefunctions at the nuclear center and decay exponentially at large distances, making them physically accurate. However, the calculation of electron repulsion integrals using STOs is computationally demanding, limiting their practical use.
To overcome this problem, Gaussian-type orbitals (GTOs) were introduced [34]

$$\phi_{GTO}(r) = e^{-\beta r^2}, \tag{1.28}$$

where β determines the shape and width of the Gaussian. Although Gaussians do not exhibit the correct cusp behavior near the nucleus, their integrals can be evaluated analytically in a much simpler and faster manner. Early basis sets, such as STO-nG, approximated each atomic orbital with n Gaussians, collectively referred to as contracted Gaussian-type orbitals (CGTOs) [35]

$$\phi_{GTOn}(r) = \sum_{i=1}^{n} D_i e^{-\beta_i r^2}, \tag{1.29}$$

where D_i refers to the contraction coefficient for GTO_i. In this approach, multiple primitive Gaussians with different exponents and coefficients are linearly combined to approximate a single STO. This method retains the computational advantages of Gaussians while mimicking the shape of STOs more closely. Figure 1.4 illustrates the differences between a single STO function, a single Gaussian function, and three prim-

itive Gaussians whose sum forms a contracted Gaussian function that more closely resembles the STO shape [36]. As computational resources improved, researchers introduced split-valence basis sets like 3-21G and 6-31G, allowing for better flexibility in valence orbitals [37]. The need for accurate descriptions of chemical bonding and molecular interactions led to the introduction of polarization functions, which enable orbitals to distort under molecular interactions, and diffuse functions, which are crucial for describing weakly bound electrons in anions and for excited states [38]. More sophisticated basis sets, such as the correlation-consistent series (cc-pVDZ, cc-pVTZ, etc.), systematically improve accuracy by incorporating correlation effects. These basis sets allow chemists to balance computational efficiency and accuracy, making them indispensable tools in quantum chemistry [39].

The choice of basis function depends on a trade-off between computational cost and accuracy. Minimal basis sets like STO-3G are fast but not very accurate while split-valence basis sets 6-31G have reasonable accuracy with moderate computational cost. Polarized basis sets 6-31G(d) are good for chemical bonding and reaction studies and correlation-consistent basis sets like cc-pVTZ or cc-pVQZ are required for high-precision correlated calculations.

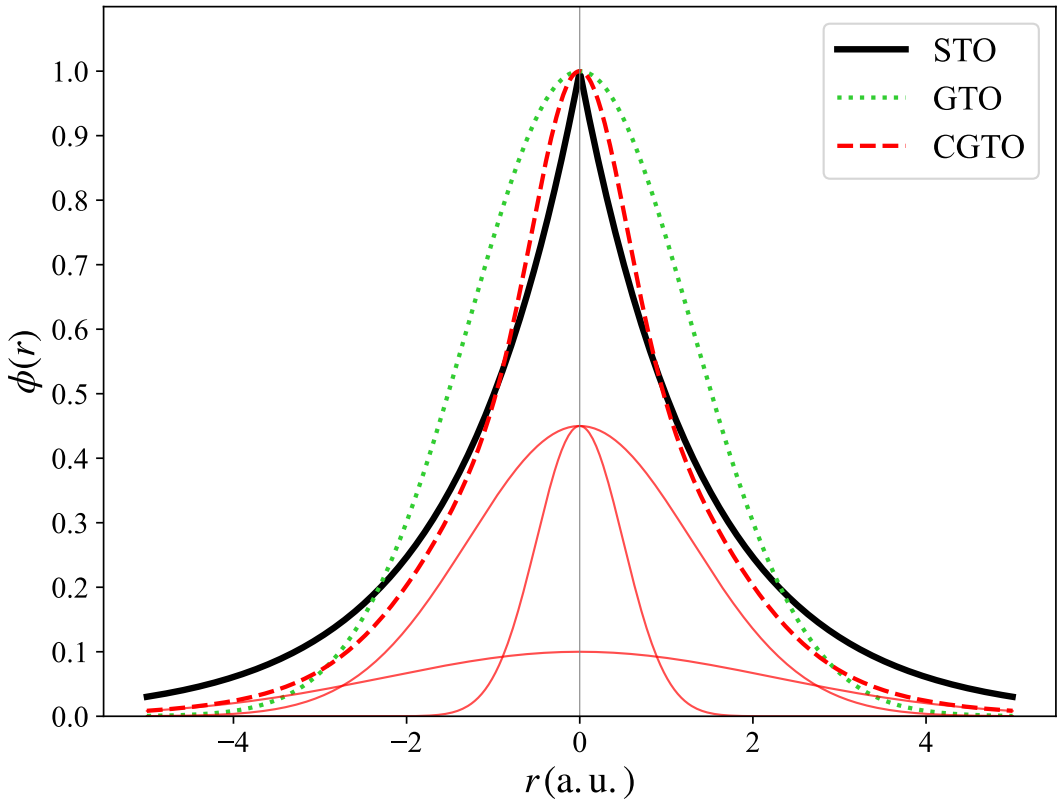


Figure 1.4: Comparison of basis functions used in quantum chemistry. The black curve shows an STO. The green dotted curve represents a single GTO. The red curves show three primitive Gaussians with different widths, and their sum (dashed red line) forms a CGTO that approximates the shape of the STO more closely.

1.2.3.9 Second Quantization

As discussed, the wavefunction of a system of N electrons can be expressed as Eq. 1.19. This formulation quickly becomes cumbersome as the number of particles increases. Second quantization simplifies this by describing quantum states in terms of occupation numbers in a predefined set of basis orbitals [40]. Instead of explicitly tracking individual electron coordinates, we use Fock space, where each basis orbital is either occupied or unoccupied. Second quantization is a reformulation of quantum mechanics in terms of creation and annihilation operators, providing a systematic way to describe many-electron states and interactions [41]. The central objects in the second quantization are the creation operator $\hat{a}_p^\dagger$ which adds an electron to orbital p

and annihilation operator $\hat{a}_p$ which removes an electron from orbital p [30].

These operators obey fermionic anticommutation relations:

$$\{\hat{a}_p, \hat{a}_q^\dagger\} = \hat{a}_p\hat{a}_q^\dagger + \hat{a}_q^\dagger\hat{a}_p = \delta_{pq}, \tag{1.30}$$

$$\{\hat{a}_p, \hat{a}_q\} = \{\hat{a}_p^\dagger, \hat{a}_q^\dagger\} = 0. \tag{1.31}$$

The many-electron wavefunction is represented as a Fock state, which specifies the occupation number of each orbital:

$$|n_1, n_2, \ldots, n_M\rangle, \tag{1.32}$$

where $n_p = 0$ or 1 for fermions. The vacuum state, representing an empty system, is:

$$|0\rangle. \tag{1.33}$$

Applying the creation operator populates an orbital:

$$\hat{a}_p^\dagger|0\rangle = |0, 0, \ldots, 1_p, \ldots, 0\rangle. \tag{1.34}$$

Similarly, applying the annihilation operator removes an electron:

$$\hat{a}_p|\ldots, 1_p, \ldots\rangle = |\ldots, 0_p, \ldots\rangle \tag{1.35}$$

or annihilate the state if the electron is not present

$$\hat{a}_p|\ldots, 0_p, \ldots\rangle = 0 \tag{1.36}$$

The electronic Hamiltonian in second quantization is written as:

$$\hat{H} = \sum_{pq} h_{pq} \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{pqrs} g_{pqrs} \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_s \hat{a}_r, \tag{1.37}$$

where h_{pq} are one-electron integrals:

$$h_{pq} = \int \psi_p^*(\mathbf{r}) \left(-\frac{\hbar^2}{2m} \nabla^2 - \sum_I \frac{Z_I e^2}{|\mathbf{r} - \mathbf{R}_I|} \right) \psi_q(\mathbf{r}) d\mathbf{r}. \tag{1.38}$$

and g_{pqrs} are two-electron integrals:

$$g_{pqrs} = \int \int \psi_p^*(\mathbf{r}_1) \psi_q^*(\mathbf{r}_2) \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_r(\mathbf{r}_2) \psi_s(\mathbf{r}_1) d\mathbf{r}_1 d\mathbf{r}_2. \tag{1.39}$$

The second quantization formalism provides a convenient framework for describing quantum states in complex many-electron systems.

1.2.3.10 Wick's Theorem

Wick's theorem reduces the complexity of computing multi-operator products by breaking down a complex product into a sum over pairwise contractions [42]. In the context of mean field theory, certain contractions vanish because of the properties of the underlying state. For HF states, this means that all higher-order correlation functions like

$$\Gamma_{\alpha\beta}^{\gamma\delta} = \langle \mathrm{HF} | a_\alpha^\dagger a_\beta^\dagger a_\delta a_\gamma | \mathrm{HF} \rangle \tag{1.40}$$

can be expressed directly in terms of the one-particle contractions, without the need to consider explicit many-body correlations.

$$\Gamma_{\alpha\beta}^{\gamma\delta} = \langle \mathrm{HF} | a_\alpha^\dagger a_\gamma | \mathrm{HF} \rangle \langle \mathrm{HF} | a_\beta^\dagger a_\delta | \mathrm{HF} \rangle - \langle \mathrm{HF} | a_\alpha^\dagger a_\delta | \mathrm{HF} \rangle \langle \mathrm{HF} | a_\beta^\dagger a_\gamma | \mathrm{HF} \rangle \tag{1.41}$$

This represents all the unique pairings between one creation and one annihilation

operator that preserve fermion number and respect anticommutation rules. In terms of the one-particle density matrix ρ, this reads:

$$\Gamma^{\gamma\delta}_{\alpha\beta} = \rho^{\gamma}_{\alpha}\rho^{\delta}_{\beta} - (\rho^{\delta}_{\alpha}\rho^{\gamma}_{\beta}) \tag{1.42}$$

An important feature of Wick's theorem is that certain contractions vanish identically due to the nature of the reference state. For example, contractions involving two creation operators or two annihilation operators are zero in the HF vacuum. This is because the HF state conserves particle number and is constructed such that only pairings between a creation and an annihilation operator yield non-zero expectation values. Thus, when we expand the four-operator product using Wick's theorem, only the nonvanishing contractions contribute to the final result, which simplifies the calculation considerably.

1.2.3.11 Density Functional Theory

Density Functional Theory (DFT) is one of the most widely used electronic structure methods in quantum chemistry. Unlike the HF method, which is a many-electron wavefunction theory, DFT reformulates the problem in terms of the electron density $\rho(\mathbf{r})$. This provides an efficient and computationally feasible framework for studying molecular properties, chemical reactivity, and solid-state materials [43].

The foundation of DFT is based on two theorems established by Hohenberg and Kohn. First, the ground-state electron density $\rho(\mathbf{r})$ uniquely determines the external potential $V_{\text{ext}}(\mathbf{r})$, and consequently, all properties of the system. Second, a universal functional for the energy, $E[\rho]$, exists, and the true ground-state density minimizes this functional [44].

These theorems imply that solving the electronic Schrödinger equation can be reformulated as a minimization problem over the electron density. Since the exact functional $E[\rho]$ is unknown, Kohn and Sham introduced a practical approach by map-

ping the interacting many-electron system onto a system of non-interacting electrons moving in an effective potential. The total energy functional is written as:

$$E[\rho] = T_s[\rho] + E_{\text{ext}}[\rho] + E_{\text{H}}[\rho] + E_{\text{XC}}[\rho], \tag{1.43}$$

where $T_s[\rho]$ is the kinetic energy of non-interacting electrons, $E_{\text{ext}}[\rho]$ is the external potential energy due to nuclei, $E_{\text{H}}[\rho]$ is the classical Coulomb energy (Hartree term), $E_{\text{XC}}[\rho]$ is the exchange-correlation energy, which includes all quantum many-body effects [45].

The Kohn-Sham equations for the single-particle orbitals $\psi_i(\mathbf{r})$ take the form:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V_{\text{eff}}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}), \tag{1.44}$$

where the effective potential is given by:

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{XC}}(\mathbf{r}). \tag{1.45}$$

These equations are solved self-consistently in an iterative process similar to Hartree-Fock.

The KS orbitals $\psi_i(\mathbf{r})$ are single-electron wavefunctions used to construct the total electron density:

$$\rho(\mathbf{r}) = \sum_i f_i|\psi_i(\mathbf{r})|^2, \tag{1.46}$$

where f_i represents the occupation numbers. The KS wavefunction is a Slater determinant:

$$\Psi_{\text{KS}}(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \psi_1(\mathbf{r}_1) & \psi_2(\mathbf{r}_1) & \ldots & \psi_N(\mathbf{r}_1) \\ \psi_1(\mathbf{r}_2) & \psi_2(\mathbf{r}_2) & \ldots & \psi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(\mathbf{r}_N) & \psi_2(\mathbf{r}_N) & \ldots & \psi_N(\mathbf{r}_N) \end{vmatrix}. \tag{1.47}$$

Although this wavefunction formally resembles the Hartree-Fock determinant, it does not include explicit electron-electron interactions beyond the exchange-correlation potential.

The accuracy of DFT depends on the approximation used for the exchange-correlation functional $E_{\mathrm{XC}}[\rho]$. Several approximations exist:

- **Local Density Approximation (LDA)**: Assumes $E_{\mathrm{XC}}[\rho]$ depends only on the local electron density. This is generally a good assumption for solids.
- **Generalized Gradient Approximation (GGA)**: Incorporates the density gradient $\nabla\rho(\mathbf{r})$ to improve accuracy, like PBE and BLYP functionals.
- **Meta-GGA Functionals**: Extend GGA by including the kinetic energy density $\tau(\mathbf{r})$ for further flexibility, e.g., M06-L.
- **Range-Separated and Dispersion-Corrected Functionals**: Improve long-range interactions and van der Waals effects. Examples include ωB97M-V and DFT-D corrections.

1.2.3.12 Quantum Theory of Atoms in Molecules

The Quantum Theory of Atoms in Molecules (QTAIM) provides a framework for analyzing electronic structure by partitioning molecular space based on the topology of the electron density [46, 47, 48]. The spatial distribution of density, $\rho(\mathbf{r})$, is dominated by the nuclei, which act as strong attractors of electron density. The regions associated with individual atoms in a molecule is defined by a zero-flux surface, which is determined by the condition

$$\nabla\rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r}) = 0, \tag{1.48}$$

where $\nabla\rho(\mathbf{r})$ is the gradient of the electron density $\rho(\mathbf{r})$, and $\mathbf{n}(\mathbf{r})$ is the unit normal vector to the zero-flux surface that separates atomic basins. This condition ensures that electrons within an atomic basin are primarily influenced by their respective nucleus, forming a quantum mechanical definition of an atom in a molecule (AIM). QTAIM characterizes molecular structure using bond critical points (BPCs) in the electron density, where the first derivative of $\rho(\mathbf{r})$ vanishes:

$$\nabla\rho(\mathbf{r}) = \mathbf{0}. \tag{1.49}$$

and a set of bond paths is an indicator of bond strength, characterized by the Laplacian of the electron density ($\nabla^2\rho_{\mathrm{BCP}}$):

- $\nabla^2\rho_{\mathrm{BCP}} < 0$: Covalent bonding (charge accumulation).
- $\nabla^2\rho_{\mathrm{BCP}} > 0$: Ionic, hydrogen bonding, or van der Waals interactions, etc. (close-shell interactions)

This framework enables the construction of larger systems from smaller fragments, as demonstrated by Chang and Bader [49], who showed that polypeptides can be assembled from transferable functional groups with minimal loss of accuracy. QTAIM correlation matrices can also serve as predictive descriptors of molecular properties, supporting the idea that matrix representations of electronic structure capture chemically meaningful information [50, 51, 52].

1.2.3.13 Transferable Atom Equivalent

The Transferable Atom Equivalent (TAE) method is grounded in Bader's AIM theory. Within this framework, molecular properties are reconstructed using libraries of precomputed atomic fragments obtained from reference quantum chemical calculations [53, 54]. Each atomic fragment is characterized by its chemical environment, enabling its transferability across different molecular systems.

Molecular reconstruction is carried out using the RECON algorithm, which selects and positions appropriate atomic equivalents based on chemical descriptors such as connectivity, hybridization, and local topology. The selected atomic densities are then aligned within the global molecular framework using quaternion-based fitting procedures, ensuring proper orientation and spatial consistency with the target structure. Following assembly, the electron density may be refined iteratively to account for environment-specific effects, thereby improving the accuracy of the reconstructed molecular properties while maintaining computational efficiency. TAE-based descriptors have been successfully applied in QSPR and QSRR studies for large biomolecules, where they provide efficient approximations of electronic properties for predictive modeling tasks [55].

1.2.3.14 Fragment Molecular Orbital

The Fragment Molecular Orbital (FMO) method is an approximate quantum mechanical approach designed to enable electronic structure calculations for large molecular systems, such as proteins, by partitioning them into smaller, computationally tractable fragments [56, 57]. Within the FMO framework, a molecular system is partitioned into N fragments. The electronic structure of each fragment is computed using a fragment Hamiltonian (H_I), which includes both internal contributions and the electrostatic potential arising from all other fragments in the system. The fragment Hamiltonian is expressed as

$$H_I = \sum_{i}^{n_I} \left(-\frac{1}{2}\nabla_i^2 - \sum_{s}^{\text{all atoms}} \frac{Z_s}{r_{si}} + \sum_{J \neq I}^{N} \int \frac{\rho_J(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|} d\mathbf{r}' \right) + \sum_{i>j}^{n_I} \frac{1}{r_{ij}}, \tag{1.50}$$

where n_I is the number of electrons in fragment I, Z_s is the nuclear charge of atom s, and $\rho_J(\mathbf{r}')$ denotes the electron density of fragment J. The total energy is given by

$$E = \sum_{I<J}^{N} E_{IJ} - (N-2)\sum_{I=1}^{N} E_I + \sum_{s<t}^{\text{all atoms}} \frac{Z_s Z_t}{r_{st}}, \tag{1.51}$$

where E_I and E_{IJ} represent the electronic energies of individual fragments and fragment pairs, respectively, and the final term accounts for nuclear repulsion.

1.2.3.15 Kernel Energy Method

The Kernel Energy Method (KEM) is a fragmentation-based quantum mechanical approach that enables efficient calculation of electronic properties in large molecular systems by decomposing them into smaller units, referred to as kernels [58, 59, 60, 61, 62, 63, 64]. Based on a many-body expansion, KEM replaces a full-system *ab initio* calculation with a series of calculations on reduced subsystems, significantly lowering computational cost while maintaining reasonable accuracy. This makes it particularly useful for studying biomolecules and extended systems.

In KEM, a molecular property P, such as the total energy, is approximated using an inclusion–exclusion scheme. In the second-order approximation (KEM2) [65, 66], the property is expressed as

$$P_{KEM2} = \sum_{a=1}^{m-1}\sum_{b=a+1}^{m} P_{ab} - (m-2)\sum_{c=1}^{m} P_c \tag{1.52}$$

where m is the total number of kernels, P_{ab} is the property calculated for the double kernel (the union of kernels a and b) and P_c is the property for the single kernel c. If only chemically bonded double kernels are considered for a linear chain, the approximation simplifies to summing adjacent pairs and subtracting the internal single kernels. The system is partitioned so that each atom belongs to a single kernel, and covalent bonds between kernels are capped with hydrogen atoms to preserve chemical

structure.

1.2.3.16 Energy Decomposition Analysis

Energy Decomposition Analysis (EDA) is a computational technique used to analyze and interpret intermolecular interactions in computational chemistry [67, 68]. Given that electronic structure methods such as HF or DFT provide total interaction energies without direct insight into their components, EDA offers a way to break down these interactions into physically meaningful terms.

The total interaction energy ΔE_{tot} between two fragments (molecules, ligands, or functional groups) is computed as:

$$\Delta E_{\text{tot}} = E_{\text{complex}} - (E_{\text{fragment 1}} + E_{\text{fragment 2}}), \tag{1.53}$$

where E_{complex} is the energy of the bound system, $E_{\text{fragment 1}}$ and $E_{\text{fragment 2}}$ are the energies of the individual fragments.

1.2.3.17 Natural Bond Orbital Analysis

Natural Bond Orbital (NBO) analysis is an insightful method for interpreting molecular electronic structure by transforming delocalized molecular orbitals into a set of localized orbitals that provide a chemically intuitive picture of bonding [69]. The transformation from delocalized MOs ψ_i to localized NBOs $\phi_{\text{NBO},i}$ is given by:

$$\phi_{\text{NBO},i} = \sum_j C_{ij}\psi_j, \tag{1.54}$$

where C_{ij} are transformation coefficients that maximize electron localization.

1.2.3.18 Size Extensiveness

The HF method is size-extensive but lacks electron correlation [70]. Size extensivity is an essential property for correctly modeling quantum chemical systems [71]. This property ensures the total energy of a system scales correctly with the number of particles. A method is size-consistent if the computed energy of a system consisting of two non-interacting subsystems (A and B) satisfies:

$$E(A+B) = E(A) + E(B). \tag{1.55}$$

1.2.3.19 Post-Hartree-Fock Methods

The post-Hartree-Fock (post-HF) methods are methods that improve upon HF theory by incorporating electron correlation [30, 32]. These methods provide more accurate solutions to the electronic Schrödinger equation and are essential for predicting molecular properties with high precision. They fall into several categories.

1.2.3.20 Configuration Interaction

Configuration Interaction (CI) improves upon HF by expressing the wavefunction as a linear combination of Slater determinants [30], accounting for electronic excitations:

$$\Psi_{\text{CI}} = c_0 \Psi_{\text{HF}} + \sum_{ia} c_i^a \Psi_i^a + \sum_{ijab} c_{ij}^{ab} \Psi_{ij}^{ab} + \dots \tag{1.56}$$

where Ψ_{HF} is the reference Hartree-Fock determinant, Ψ_i^a represents a single excitation, where an electron is promoted from occupied orbital i to virtual orbital a, Ψ_{ij}^{ab} represents a double excitation, and so on. Different levels of CI approximations exist, such as CIS (CI Singles), which only includes single excitations, CISD (CI Singles and Doubles), and so on. The exact solution for the electronic Schrödinger equation within a given basis set can be obtained by Full CI (FCI), which includes all possible

excitations but is computationally intractable for all but the smallest systems.

1.2.3.21 Møller-Plesset Perturbation Theory

Møller-Plesset Perturbation Theory (MPn) treats electron correlation as a perturbation to the HF wavefunction [72]. The total energy is expressed as a series expansion:

$$E_{\mathrm{MP}} = E_{\mathrm{HF}} + E^{(2)} + E^{(3)} + E^{(4)} + \ldots \tag{1.57}$$

where $E^{(n)}$ represents the correction at the n-th order. A common MPn method is MP2 (second-order perturbation theory) which is widely used due to its balance between accuracy and computational cost.

1.2.3.22 Multi Reference Methods

For describing systems where a single reference HF wavefunction is insufficient, such as bond breaking in any chemical system or transition metal complexes, multireference methods are essential [73]. These methods explicitly account for strong electron correlation. One of the most widely used multireference methods is the Complete Active Space Self Consistent Field (CASSCF), where an active space of important orbitals is selected, and full electron correlation is included within this space [74]. The wavefunction is expressed as a linear combination of multiple Slater determinants:

$$|\Psi_{\mathrm{CASSCF}}\rangle = \sum_i C_i |\Phi_i\rangle$$

where Φ_i are different configurations (Slater determinants) formed within the active space.

While CASSCF captures static correlation, it lacks dynamic correlation, which can be introduced using post-CASSCF methods like Complete Active Space Second-Order Perturbation Theory (CASPT2 [75]).

1.2.3.23 Coupled Cluster Theory

Coupled Cluster (CC) theory is one of the most accurate post-HF methods [76]. Instead of a linear expansion like CI, CC uses an exponential ansatz:

$$\Psi_{\mathrm{CC}} = e^{\hat{T}}\Psi_{\mathrm{HF}}, \tag{1.58}$$

where $\hat{T}$ is the cluster operator:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots \tag{1.59}$$

Here $\hat{T}_1$ is a single excitation, $\hat{T}_2$ is a double excitation, $\hat{T}_3$, $\hat{T}_4$, etc. are higher excitation operators. One of the popular CC methods is CCSD (Coupled-Cluster Singles and Doubles), which accounts for single and double excitations. CCSD with a perturbative treatment of triple excitations is called CCSD(T), which is considered the "gold standard" in quantum chemistry [77].

Despite its accuracy, canonical CCSD(T) scales as $\mathcal{O}(N^7)$, making it computationally expensive for large systems. To address this limitation, the Domain-Based Local Pair Natural Orbital Coupled Cluster (DLPNO-CCSD(T)) method was developed [78]. It constructs an efficient representation of the virtual orbital space using Pair Natural Orbitals (PNOs) [79], where only the most important virtual orbitals contributing to electron correlation are retained. The key idea in PNO-based methods is that the virtual orbital space can be efficiently compressed using natural orbitals derived from the correlation of electron pairs. For a given pair of localized occupied orbitals (i, j), the corresponding PNOs $\{\psi_n^{ij}\}$ are obtained by diagonalizing the MP2-based pair density matrix:

$$D_{ab}^{ij} = \sum_c t_{ij}^{ac}(t_{ij}^{bc})^*, \tag{1.60}$$

where a, b, c are indices over virtual orbitals in the pair domain, and t_{ij}^{ab} are the MP2

amplitudes for exciting electrons from i, j to virtual orbitals a, b. Diagonalizing D^{ij} yields eigenvalues n_n^{ij}, which serve as occupation numbers indicating the contribution of each pair natural orbital to the correlation energy.

Orbitals with small occupation numbers n_n^{ij} are truncated based on a threshold T_{CutPNO}, retaining only the most significant contributions:

$$n_n^{ij} > T_{\text{CutPNO}}. \tag{1.61}$$

This truncation ensures that only the most important virtual orbitals for each pair are retained, dramatically reducing the computational cost of CCSD(T) while maintaining accuracy. The result is a sparse and localized representation of the wavefunction.

1.2.3.24 Density Matrix Renormalization Group

The Density Matrix Renormalization Group (DMRG) is a highly efficient numerical method for solving strongly correlated quantum systems [80]. DMRG was originally developed in the context of condensed matter physics and has since been adapted for quantum chemistry. DMRG overcomes an exponential growth in computational cost by using a more compact representation of the wavefunction, which captures only the most relevant quantum correlations [81].

A general many-body wavefunction for a system with L orbitals can be written in the basis $|n_1, n_2, \ldots, n_L\rangle$ as:

$$|\Psi\rangle = \sum_{n_1, n_2, \ldots, n_L} C_{n_1, n_2, \ldots, n_L} |n_1, n_2, \ldots, n_L\rangle. \tag{1.62}$$

The tensor C contains exponentially many coefficients. This exponential growth makes it impractical to perform exact electronic structure calculations for systems with large L. DMRG avoids this issue by factorizing C into a product of smaller

tensors using the Matrix Product State (MPS) formalism:

$$C_{n_1,\ldots,n_L} = \sum_{\{\alpha\}} A^{n_1}_{\alpha_0\alpha_1} A^{n_2}_{\alpha_1\alpha_2} \ldots A^{n_L}_{\alpha_{L-1}\alpha_L}, \tag{1.63}$$

where each A^{n_i} is a three-index tensor. The indices α_i are called bond indices or auxiliary indices and its dimension, called the m, determine the accuracy of the approximation.

The power of MPS lies in its ability to exploit the entanglement in quantum systems by using the Schmidt decomposition [82]. Given a bipartition of a quantum system into two subsystems A and B, any wavefunction can be written as:

$$|\Psi\rangle = \sum_{i=1}^{n} \lambda_i |\psi_i^A\rangle \otimes |\psi_i^B\rangle, \tag{1.64}$$

where $|\psi_i^A\rangle$ and $|\psi_i^B\rangle$ are orthonormal states in each subsystem, and λ_i are the Schmidt coefficients. The number of significant Schmidt coefficients determines the minimal number of terms needed to accurately represent the entanglement across the bipartition. In the MPS representation, this number corresponds to the bond dimension m, which sets the size of the auxiliary (virtual) indices α_i connecting neighboring tensors. A larger m enables the MPS to capture more entanglement, while a smaller m reduces computational cost at the expense of accuracy.

The Schmidt decomposition can be calculated through the singular value decomposition (SVD) [82]. The full wavefunction tensor is first reshaped into a matrix M by grouping the indices associated with a chosen bipartition of the system. This matrix is then factorized using SVD:

$$M = U\Sigma V^\dagger, \tag{1.65}$$

where U and V are unitary matrices containing the orthonormal basis vectors for subsystems A and B, respectively, and Σ is a diagonal matrix whose entries are the

singular values corresponding to the Schmidt coefficients λ_i.

Figure 1.5 illustrates the construction of the MPS through successive tensor decompositions of the full wavefunction coefficient tensor. Starting from the original tensor $C_{n_1 n_2 \ldots n_L}$, the wavefunction is systematically factorized using the SVD. At each step, a local site tensor A_i is extracted while the remaining part is passed on as a contracted core tensor. This iterative process continues until the entire wavefunction is represented as a chain of low-rank tensors. Each tensor corresponds to a single orbital and contains one physical leg, pairing upward, representing the local orbital occupation, and two virtual or bond legs that connect to neighboring tensors.

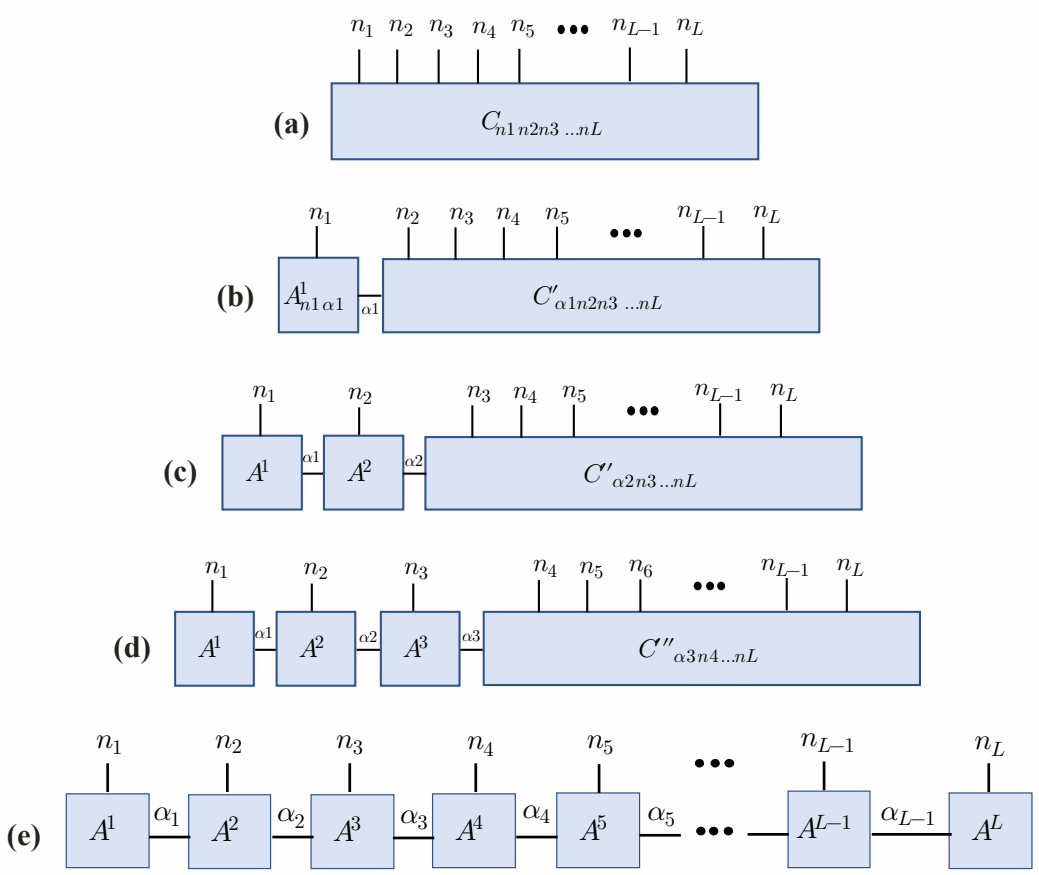


Figure 1.5: Graphical representation of an MPS tensor network with physical indices n_i and bond indices α_i. (a) The original wavefunction coefficient tensor $C_{n_1 n_2 \ldots n_L}$, which encodes all coefficients of the many-body wavefunction. (b) Factorization begins by reshaping and decomposing the tensor into a smaller tensor A_1 and a contracted core C' via SVD. (c) This process continues iteratively, decomposing the updated core into successive tensors $A_2, A_3, \ldots$, with new auxiliary indices α_i introduced at each step. (d) The final MPS structure emerges as a chain of tensors $A_1, A_2, \ldots, A_L$, each linked by shared auxiliary indices. (e) Full MPS representation of the wavefunction as a tensor network: each node corresponds to a local site tensor, and the virtual bonds encode entanglement between neighboring sites.

For the MPS, efficient computation of physical observables relies on bringing the wavefunction into a canonical form. This involves orthonormalizing the tensors in the MPS with respect to a specific site, known as the orthogonality center. The figure 1.6 illustrates this concept by showing an MPS structure centered at site L. On the left side of the orthogonality center, tensors $A_1, A_2, \ldots, A_{L-1}$ are left-orthonormal. This means that the contraction over the physical and left bond indices

satisfies the isometry condition:

$$\sum_{n_i} (A_i^{n_i})^\dagger A_i^{n_i} = I. \tag{1.66}$$

On the right-hand side of the MPS, tensors are either unnormalized or prepared for right-orthonormalization in the reverse sweep. The central tensor A_L, often referred to as the mixed-canonical tensor, is the only one not constrained by orthonormality and acts as the optimization target during DMRG sweeps. This structure is crucial for reducing the cost of evaluating expectation values such as $\langle\Psi|\hat{H}|\Psi\rangle$, since it eliminates redundant tensor contractions and focuses optimization on a local environment.

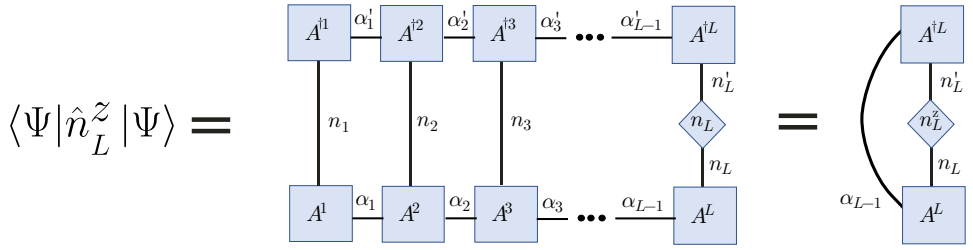


Figure 1.6: Canonical form of a MPS centered at site L. Tensors to the left of L are left-orthonormal, ensuring identity contractions over bond indices α_i'. The central tensor A_L is the orthogonality center and is not constrained by normalization. This structure is essential for efficient DMRG optimizations and operator evaluations.

DMRG optimizes the MPS wavefunction iteratively using a procedure known as the sweep algorithm:

1. Initialization: Start with a reference wavefunction, often from HF calculation.
2. Left-to-right sweep: Optimize tensors site-by-site while keeping other tensors fixed.
3. Right-to-left sweep: Reverse the optimization direction.
4. Convergence: Repeat the sweeps until the energy converges to the desired precision.

The electronic Hamiltonian within the MPS formalism will be expressed as a Matrix Product Operator (MPO)[83, 84, 85]. MPOs decompose the total Hamiltonian into a chain of local operator-valued tensors. The MPO format is particularly advantageous when dealing with local Hamiltonians acting on one-dimensional chains, which obey the area law of entanglement.

1.2.3.25 Locality in Electronic Structure

A fundamental concept underpinning the efficiency of modern electronic structure methods is the locality of quantum correlations in many-body systems. This idea is closely connected to the area law of entanglement entropy, originally developed in condensed matter physics in the study of geometric entropy in quantum field theory, motivated by black hole physics. The area law states that the entanglement between two spatial regions scales with the size of their boundary rather than their volume [86, 87], implying that quantum correlations are predominantly short-ranged and localized in space [88, 89].

Closely related to this is the nearsightedness principle introduced by Kohn, which states that local electronic properties are only weakly influenced by distant perturbations [90, 91, 92]. As a consequence, the one-particle density matrix exhibits spatial decay, allowing the electronic structure of large systems to be accurately described using local information.

Together, these principles establish the theoretical basis for fragmentation approaches and reduced representations of the electronic structure to describe electronic structure. By exploiting the locality of electronic interactions, such methods enable scalable treatments of large molecular systems.

1.2.3.26 Density Matrix Formalism

The density matrix formalism provides an alternative framework to describe quantum systems, particularly useful for systems with many interacting particles. Instead of using a wavefunction, the state of the system is represented by the density operator $\hat{\rho}$. A pure state is

$$\hat{\rho} = |\psi\rangle\langle\psi|, \tag{1.67}$$

The density operator $\hat{\rho}$ is a Hermitian, positive semi-definite and normalized $\mathrm{Tr}(\hat{\rho}) = 1$.

1.2.3.27 Reduced Density Matrices

In many-body quantum systems, we are often interested in a subset of degrees of freedom while ignoring the rest. For example, we might be interested in a subsystem A, embedded in an environment B, which we prefer to ignore. The state of such a subsystem can be described using a reduced density matrix (RDM), obtained by performing a partial trace over the remaining degrees of freedom (the environment) [93, 94, 95, 96, 97].

Suppose the full system is described by a pure state $|\Psi\rangle$ in a Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, where $\mathcal{H}_A$ and $\mathcal{H}_B$ represent the subsystems of interest and environment, respectively. The reduced density matrix of subsystem A is:

$$\hat{\rho}_A = \mathrm{Tr}_B(|\Psi\rangle\langle\Psi|), \tag{1.68}$$

where Tr_B denotes the partial trace over the degrees of freedom in B. The $\hat{\rho}_A$ contains all the information necessary to compute physical observables within A.

1.2.3.28 Quantum Entanglement

The RDM allows us to characterize entanglement. A standard measure of entanglement is the von Neumann entropy. For a pure state of a bipartite system, the von Neumann entropy can be stated as [98]:

$$S(\rho_A) = -\mathrm{Tr}(\rho_A \log \rho_A), \tag{1.69}$$

where $\rho_A = \mathrm{Tr}_B(\rho_{AB})$ is the RDM obtained by tracing out subsystem B.

1.2.3.29 One-Orbital Reduced Density Matrix

The One-Orbital Reduced Density Matrix (1-RDM), ρ_i, describes the quantum state of a single spatial orbital i when all other orbitals are traced out from the full many-body wavefunction. Each spatial orbital has a four-dimensional Hilbert space:

$$\mathcal{H}_i = \{\ |-\rangle, |\uparrow\rangle, |\downarrow\rangle, |\updownarrow\rangle\ \}, \tag{1.70}$$

representing the empty state, spin-up, spin-down, and doubly-occupied states, respectively. These four configurations form the local basis $\mathcal{H}_i$ of orbital i, so any operator acting locally on orbital i can be represented as a 4×4 matrix.

This means there are 16 linearly independent operators that span the space of all local operators on orbital i. Each operator $\mathcal{O}^{(m)} = |k\rangle\langle l|$ with $|k\rangle\langle l| \in \{\ |-\rangle, |\uparrow\rangle, |\downarrow\rangle, |\updownarrow\rangle\ \}$is a sparse matrix with a single non-zero entry of value 1 and corresponds to a specific transition between these four states.

$$\mathcal{O}^{(1)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \mathcal{O}^{(2)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, etc. \tag{1.71}$$

The complete set of these operators is summarized in Table 1.1.

Table 1.1: Complete set of single-orbital transition operators $\mathcal{O}^{(m)} = |k\rangle\langle l|$, where the row and column labels represent the final state $|k\rangle$ and initial state $|l\rangle$, respectively. Each operator acts as a projector or transition matrix in the four-dimensional local Hilbert space of a single spatial orbital, with a single non-zero element at position (k, l).

$\mathcal{O}^{(m)}$	$-$	$\downarrow$	$\uparrow$	$\updownarrow$
$-$	$\mathcal{O}^{(1)}$	$\mathcal{O}^{(2)}$	$\mathcal{O}^{(3)}$	$\mathcal{O}^{(4)}$
$\downarrow$	$\mathcal{O}^{(5)}$	$\mathcal{O}^{(6)}$	$\mathcal{O}^{(7)}$	$\mathcal{O}^{(8)}$
$\uparrow$	$\mathcal{O}^{(9)}$	$\mathcal{O}^{(10)}$	$\mathcal{O}^{(11)}$	$\mathcal{O}^{(12)}$
$\updownarrow$	$\mathcal{O}^{(13)}$	$\mathcal{O}^{(14)}$	$\mathcal{O}^{(15)}$	$\mathcal{O}^{(16)}$

The ρ_i is therefore a diagonal 4×4 Hermitian, positive semidefinite matrix. In systems that conserve particle number and spin projection, the off-diagonal elements of ρ_i vanish, and the matrix becomes diagonal in the occupation number basis. Its structure is therefore simplified to:

$$\rho_i = \begin{pmatrix} \langle \mathcal{O}_i^{(1)} \rangle & 0 & 0 & 0 \\ 0 & \langle \mathcal{O}_i^{(6)} \rangle & 0 & 0 \\ 0 & 0 & \langle \mathcal{O}_i^{(11)} \rangle & 0 \\ 0 & 0 & 0 & \langle \mathcal{O}_i^{(16)} \rangle \end{pmatrix}, \tag{1.72}$$

The eigenvalues of ρ_i satisfy:

$$\sum_{\alpha=1}^{4} w_{\alpha,i} = 1, \quad w_{\alpha,i} \geq 0. \tag{1.73}$$

The mapping between these diagonal elements and their corresponding projectors is summarized in Table 1.2.

Table 1.2: Mapping of the diagonal entries of the ρ_i to the corresponding local projector operators $\mathcal{O}_i^{(m)}$. Each numbered entry indicates the operator index m projecting onto the respective occupation basis state.

ρ_i	$\vert - \rangle$	$\vert \downarrow \rangle$	$\vert \uparrow \rangle$	$\vert \uparrow\downarrow \rangle$
$\vert - \rangle$	1			
$\vert \downarrow \rangle$		6		
$\vert \uparrow \rangle$			11	
$\vert \uparrow\downarrow \rangle$				16

The operators $\mathcal{O}_i^{(m)}$ are defined in Table 1.3 in terms of fermionic creation and annihilation operators $c_\sigma^\dagger$, c_σ, and number operators $n_\sigma = c_\sigma^\dagger c_\sigma$, where $\sigma \in \{\uparrow, \downarrow\}$. The operator set is based on the formalism introduced in Ref. [99], with corrections applied to $\mathcal{O}^{(8)}$ and $\mathcal{O}^{(14)}$.

Table 1.3: Sixteen single-orbital operators. obtained from [99], operators $\mathcal{O}^{(8)}$ and $\mathcal{O}^{(14)}$ are corrected.

$\mathcal{O}^{(1)}$	$1 - n_\uparrow - n_\downarrow + n_\uparrow n_\downarrow$
$\mathcal{O}^{(2)}$	$c_\downarrow - n_\uparrow c \downarrow$
$\mathcal{O}^{(3)}$	$c_\uparrow - n_\downarrow c_\uparrow$
$\mathcal{O}^{(4)}$	$c_\downarrow c_\uparrow$
$\mathcal{O}^{(5)}$	$c_\downarrow^\dagger - n \uparrow c_\downarrow^\dagger$
$\mathcal{O}^{(6)}$	$n_\downarrow - n \uparrow n_\downarrow$
$\mathcal{O}^{(7)}$	$c_\downarrow^\dagger c_\uparrow$
$\mathcal{O}^{(8)}$	$n_\downarrow c_\uparrow$
$\mathcal{O}^{(9)}$	$c_\uparrow^\dagger - n \downarrow c_\uparrow^\dagger$
$\mathcal{O}^{(10)}$	$c_\downarrow c_\uparrow^\dagger$
$\mathcal{O}^{(11)}$	$n_\uparrow - n_\uparrow n_\downarrow$
$\mathcal{O}^{(12)}$	$n_\uparrow c_\downarrow$
$\mathcal{O}^{(13)}$	$c_\downarrow^\dagger c_\uparrow^\dagger$
$\mathcal{O}^{(14)}$	$n_\downarrow c_\uparrow^\dagger$
$\mathcal{O}^{(15)}$	$n_\uparrow c_\downarrow^\dagger$
$\mathcal{O}^{(16)}$	$n_\uparrow n_\downarrow$

1.2.3.30 Two-Orbital Reduced Density Matrix

The two-orbital reduced density matrix (2-RDM), ρ_{ij}, contains all quantum mechanical information about the joint state of orbitals i and j. Each orbital contributes a four-dimensional Hilbert space, so the joint Hilbert space for the pair is the tensor product:

$$\mathcal{H}_{ij} = \mathcal{H}_i \otimes \mathcal{H}_j, \tag{1.74}$$

which is 16-dimensional. Consequently, ρ_{ij} is a 16×16 Hermitian, positive semidefinite matrix. A particularly important quantity derived from the 2-RDM is the two-orbital

entropy $s(2)_{ij}$, defined as:

$$s(2)_{ij} = -\sum_{\alpha=1}^{16} w_{\alpha,ij} \ln w_{\alpha,ij}, \tag{1.75}$$

which measures how entangled the pair $\{i, j\}$ is with the rest of the system. A higher entropy indicates greater entanglement between the selected orbitals and the environment.

The structure of the 2-RDM in this basis is provided in Table 1.4. Each matrix element represents a quantum transition between an initial and a final occupation configuration of the orbital pair (i, j). The notation m/n is used to label each matrix element according to the corresponding local operators acting on orbitals i and j, respectively.

Table 1.4: 2-RDM in the occupation number basis. The row and column labels indicate the final and initial states, respectively. Each entry corresponds to a transition matrix element between two configurations of orbitals i and j, labeled by symbols for empty (—), spin-up (↑), spin-down (↓), and doubly occupied (⇵) states.

ρ_{ij}	— —	— ↓	↓ —	— ↑	↑ —	↓ ↓	— ⇵	↓ ↑	↑ ↓	⇵ —	↑ ↑	↓ ⇵	⇵ ↓	↑ ⇵	⇵ ↑	⇵ ⇵
— —	1/1															
— ↓		1/6	2/5													
↓ —		5/2	6/1													
— ↑				1/11	3/9											
↑ —				9/3	11/1											
↓ ↓						6/6										
— ⇵							1/16	2/15	3/14	4/13						
↓ ↑							5/12	6/ 11	7/10	8/9						
↑ ↓							9/8	10/7	11/6	12/5						
⇵ —							13/4	14/3	15/2	16/1						
↑ ↑											11/11					
↓ ⇵												6/16	8/14			
⇵ ↓												14/8	16/6			
↑ ⇵														11/16	12/15	
⇵ ↑														15/12	16/11	
⇵ ⇵																16/16

1.2.3.31 Mutual Information

Mutual information (MI) is a central concept from information theory that captures both classical and quantum correlations, making it a powerful diagnostic tool for analyzing electronic structure in correlated wavefunctions. In classical information theory, MI between two random variables X and Y is defined as:

$$I(X;Y) = \sum_{x,y} P(x,y) \log \left(\frac{P(x,y)}{P(x)P(y)} \right), \tag{1.76}$$

where $P(x,y)$ is the joint probability distribution. This expression quantifies the amount of information shared between X and Y. The classical MI reflects only statistical (classical) correlations and does not account for quantum entanglement.
In contrast, quantum MI arises due to entanglement, meaning the state of one orbital cannot be described independently of another.
The total correlation between two orbitals i and j can be captured by

$$I_{ij} = [s(2)_{ij} - s(1)_i - s(1)_j]\,(1 - \delta_{ij}). \tag{1.77}$$

This includes both classical correlations (due to probabilistic mixtures) and quantum correlations (due to entanglement).
Unlike $s(2)_{ij}$, which measures the entanglement between orbitals i, j and the environment, I_{ij} removes environmental effects, isolating direct correlations between i and j. This allows for a refined analysis of orbital-pair interactions. Figure 2.1 shows schematic representation of 1-, 2-orbital RDMs and MI in Glycine-Glycine dipeptide.

1.2.3.32 Atomic Mutual Information

Recent developments in quantum information theory (QTI) have underscored the importance of spatially localized entanglement in molecular systems. Ding et al. [100] emphasized that, to make entanglement operationally meaningful and chemically

interpretable, orbitals must be constructed to be localized on as few atomic centers as possible. This spatial localization ensures that entanglement can be harnessed through local operations and reflects physical separability within a molecule.

Building on this perspective, we transition from the orbital to the atomic level by introducing Atomic Mutual Information (AMI). To achieve this, we employ orthogonalized atomic orbitals (section 1.2.3.35) that are strictly localized on single atomic centers. For atoms A and B, the AMI is approximated by summing over mutual information between all orbitals $i \in A$ and $j \in B$:

$$I_{AB} \approx \sum_{i=1}^{L_A} \sum_{j=1}^{L_B} I_{ij}, \tag{1.78}$$

where L_A and L_B are the number of orbitals associated with atoms A and B, respectively. This coarse-grained measure is computationally more feasible and aligns better with chemical intuition.

While this approximation is practical and useful, it does not capture the full many-body correlations between atoms. The exact definition of AMI involves computing the von Neumann entropies of the RDMs for whole atoms A, B, and its pair AB:

$$I_{AB} = S(\rho_A) + S(\rho_B) - S(\rho_{AB}), \tag{1.79}$$

where ρ_A and ρ_B are the RDMs of atoms A and B, and ρ_{AB} is the joint RDM. However, the exact AMI is computationally demanding because constructing atomic RDMs requires tracing out the rest of the system while retaining all degrees of freedom on atoms A and B, which becomes prohibitively expensive in large systems. A cartoon representation of AMI is illustrated in Figure 2.2, where atom A is connected to atom B through AMI.

1.2.3.33 Fragment Mutual Information

While the AMI provides a chemically intuitive view of correlation at the atomic level, it can be generalized to larger chemically or spatially defined regions. This leads to the concept of Fragment Mutual Information (FMI), which quantifies the total correlation between arbitrary fragments of the system, such as functional groups, ligands, or residues in a biomolecule.

For two fragments F and G, consisting of sets of orbitals $\{i \in F\}$ and $\{j \in G\}$, respectively, the FMI is defined analogously to AMI by summing over all pairwise orbital mutual information terms:

$$I_{FG} \approx \sum_{i=1}^{L_F} \sum_{j=1}^{L_G} I_{ij}. \tag{1.80}$$

This quantity captures the overall strength of correlation between two fragments, making it a valuable tool in analyzing chemical interactions. Like AMI, FMI provides a more coarse-grained view of the correlation network, and it is especially useful for identifying dominant interacting fragments in large-scale systems. This has applications in:

- Local correlation methods and active space selection,
- Fragmentation schemes for quantum embedding,
- Analyzing protein-ligand interactions,
- Analyzing protein folding.

FMI bridges the gap between microscopic orbital-level correlations and mesoscopic chemically meaningful interactions, thus enriching the interpretability of quantum information analyses in complex systems.

1.2.3.34 Localized Molecular Orbitals

Canonical MOs are typically delocalized across the molecule. While suitable for energy calculations, these delocalized orbitals often obscure chemical insight, such as the identification of bonds, lone pairs, and functional groups. To overcome this limitation, localized molecular orbitals (LMOs) are introduced. LMOs provide a spatially compact representation of electronic structure, enabling more intuitive chemical interpretation and facilitating methods that rely on locality, such as fragment-based electron correlation.

Orbital localization is achieved by applying a unitary transformation to the occupied canonical MOs:

$$\psi_i' = \sum_j U_{ji}\psi_j, \tag{1.81}$$

where ψ_i' are the localized orbitals and $\mathbf{U}$ is a unitary matrix chosen to optimize a localization criterion. The general form of the localization functional is often an expectation value of a two-electron operator Ω , written as:

$$\langle\Omega\rangle = \sum_{i=1}^{N_{\text{occ}}} \langle\psi_i'\psi_i'|\Omega|\psi_i'\psi_i'\rangle. \tag{1.82}$$

Different localization schemes define Ω differently, depending on whether spatial compactness or atomic character is being optimized.

The Foster–Boys (FB) [101] method seeks to minimize the spatial extent of each orbital. It does so by minimizing the variance of the electron density around its centroid. The functional to minimize is:

$$\langle\Omega\rangle_{\text{FB}} = \sum_i \langle\psi_i'|(\mathbf{r} - \langle\psi_i'|\mathbf{r}|\psi_i'\rangle)^2|\psi_i'\rangle. \tag{1.83}$$

This expression quantifies how far the orbital density extends from its centroid, and minimizing it leads to orbitals that are maximally localized in space. Optimization

is carried out via iterative methods such as Jacobi rotations or gradient descent, preserving orthonormality.

The Høyvik and Jørgensen [102] extended the FB localization functional to include higher-order central moments (i.e., the spread of each orbital around its centroid) of the orbital density.

$$\langle \Omega' \rangle_{\mathrm{HJ}} = \sum_{i=1}^{N_{\mathrm{orb}}} \left\langle \phi_i' \left| (\mathbf{r} - \langle \phi_i' | \mathbf{r} | \phi_i' \rangle)^n \right| \phi_i' \right\rangle^m \tag{1.84}$$

For $n = 2$ and $m = 1$, the original Foster–Boys criterion is recovered. Larger values of n increasingly penalize the tails of the orbital distributions, and m reduces the spread of the least compact LMO, resulting in LMOs that are more compact and tightly centered around their respective centroids.

The Edmiston–Ruedenberg (ER) [103] method maximizes the self-repulsion energy by using the inverse interelectronic distance:

$$\langle \Omega \rangle_{\mathrm{ER}} = \sum_{i=1}^{N_{\mathrm{orb}}} \left\langle \phi_i' \phi_i' \left| \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right| \phi_i' \phi_i' \right\rangle \tag{1.85}$$

Equivalently, it minimizes inter-orbital Coulomb repulsion:

$$\langle \Omega' \rangle_{\mathrm{ER}} = \sum_{i,j=1 i \neq j}^{N_{\mathrm{orb}}} \left\langle \phi_i' \phi_j' \left| \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \right| \phi_i' \phi_j' \right\rangle \tag{1.86}$$

The Pipek–Mezey (PM) [104] method localizes orbitals by maximizing their atomic character, quantified through Mulliken population analysis. The partial charge Q_A on atom A due to orbital ψ_i' is calculated using:

$$Q_A = Z_A - \sum_{i=1}^{N_{\mathrm{occ}}} \sum_{\alpha \in A}^{M_{\mathrm{basis}}} \sum_{\beta=1}^{M_{\mathrm{basis}}} c_{\alpha i} c_{\beta i} S_{\alpha\beta}, \tag{1.87}$$

where $c_{\alpha i}$ are MO coefficients, and $S_{\alpha\beta}$ are elements of the AO overlap matrix. The PM functional to be maximized is the sum of squared orbital contributions to atomic

charges:

$$\langle \Omega \rangle_{\mathrm{PM}} = \sum_{A} \sum_{i} (q_i^A)^2, \tag{1.88}$$

where q_i^A is the Mulliken population of orbital i on atom A. Maximizing this quantity leads to orbitals that are localized on individual atoms or atom pairs. Unlike Foster–Boys, PM localization does not require calculation of position operators, making it suitable for both molecular and periodic systems.

Both methods yield chemically interpretable orbitals and serve as crucial tools in wavefunction analysis and post-Hartree–Fock computations.

1.2.3.35 Canonical Orthogonalization

Despite the spatial compactness of LMOs, they are not strictly centered on a single atom, even for the most localized occupied orbitals. Moreover, LMO construction involves arbitrary choices, such as the specific localization criterion (e.g., FB, ER, PM, etc.), which can result in different localized orbital sets for the same molecular system. This inherent arbitrariness complicates their use in analyses requiring atomic resolution.

To overcome these limitations, we adopt the AO basis as a more suitable alternative for atom-centered analysis. AOs are naturally defined on individual nuclei and provide a consistent reference frame for atom-specific analysis. This choice is particularly well-motivated for AMI measures which depend on the ability to assign orbitals to atomic subspaces.

One practical issue with the AO basis is its nonorthogonality due to their spatial overlap, and this nonorthogonality introduces complications in constructing RDMs such as the trace of the 1-RDM, which should equal the number of electrons, is not well-defined unless corrected by the overlap matrix. Similar issues arise when projecting wavefunctions or performing entropic analyses, as these rely on an orthonormal basis.

To address this, we employed a straightforward and robust approach called canonical orthogonalization. Canonical orthogonalization allows for projecting MOs, which are typically expressed in a delocalized orthonormal basis, into the AO basis. This transformation facilitates the assignment of MOs to atoms by allowing the analysis of MO contributions in terms of atomic-centered functions. This scheme transforms a set of nonorthogonal AOs $\{a^\dagger_{i\sigma}\}$ into an orthonormal set $\{c^\dagger_{j\sigma}\}$ through the relation:

$$c^\dagger_{j\sigma} = \sum_{i=1}^{L} S^{-1/2}_{ij}\, a^\dagger_{i\sigma}, \tag{1.89}$$

where S is the AO overlap matrix, and $S^{-1/2}$ denotes its inverse square root obtained via eigenvalue decomposition. This transformation ensures that the resulting basis functions satisfy orthonormality by construction, while remaining sufficiently close in shape and locality to the original AOs.

This orthogonalization step is particularly crucial for evaluating AMI and FMI, where accurate partitioning of the wavefunction into atomic subspaces depends on the orthonormality of the basis. Other localization procedures can be explored as alternatives, provided they also yield orthonormal orbitals; however, canonical orthogonalization offers a simple and general solution that is well-suited for our analysis.

1.3 Outline of the Thesis

The theoretical concepts and methodological tools introduced in this chapter establish the foundation for the developments presented in the remainder of this thesis. In particular, the formulation of mutual information within a quantum information theory and the construction of atom-centered orthonormal bases provide the necessary ingredients for defining and evaluating electronic correlations at atomic and fragment levels in molecular systems.

Chapter 2 builds directly on this framework by introducing the development of AMI

as a novel tool for quantifying electronic correlations in molecular systems. This chapter focuses on the formal definition of AMI, its computational implementation, and its application to dipeptide systems. The results of this work have been published as a preprint on ChemRxiv and submitted to the Journal of Theoretical and Computational Chemistry under Manuscript ID ct-2024-016809.R1. I served as the principal author of this study, with primary responsibility for method development, algorithm design, software implementation, generation and analysis of results, and preparation of the manuscript.

In Chapter 3, the AMI framework is extended to FMI. This chapter investigates the relationship between MI correlation measures and conventional energetic descriptors, including atomic energy contributions and bonding energies. In addition, the evolution of AMI and FMI during MD simulations is examined. The results of this work have been published as a preprint on ChemRxiv and submitted to ACS Omega under Manuscript ID JP-2025-05628N. I was the principal author and led the method development, algorithm formulation, coding, numerical analysis, and manuscript preparation.

Chapter 4 introduces an innovative cut-wise reconstruction strategy for investigating AMI in full protein systems. In this approach, a full protein is divided into smaller overlapping fragments, analyzed individually, and subsequently stitched together to reconstruct electronic correlation patterns at the full protein scale. This methodology enables quantum information analysis of systems that are otherwise computationally prohibitive. The results of this work have been published in Chemical Communications. I served as the principal author and was responsible for method development, generation and analysis of results, and preparation of the manuscript.

Finally, Chapter 5 provides a comprehensive discussion of the results presented in this thesis. The findings are placed in the broader context of related studies from other academic institutes, highlighting both conceptual connections and methodological

distinctions. The chapter concludes with a summary of the main contributions of this work and an outlook on future directions for applying quantum information theory to electronic structure analysis of biomolecular systems.

Bibliography

[1] G. Barrett, Chemistry and Biochemistry of the Amino Acids, Springer Science & Business Media, 2012.

[2] A. B. Hughes, Amino Acids, Peptides and Proteins in Organic Chemistry, Protection Reactions, Medicinal Chemistry, Combinatorial Synthesis, John Wiley & Sons, 2011.

[3] D. Whitford, Proteins: Structure and Function, John Wiley & Sons, 2013.

[4] B. Rupp, Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology, Garland Science, 2009.

[5] K. Wüthrich, NMR With Proteins and Nucleic Acids, Europhys. News 17 (1) (1986) 11–13.

[6] W. Kühlbrandt, The Resolution Revolution, Sci. 343 (6178) (2014) 1443–1444.

[7] C. B. Anfinsen, Principles That Govern the Folding of Protein Chains, Sci. 181 (4096) (1973) 223–230.

[8] K. A. Dill, J. L. MacCallum, The Protein-Folding Problem, 50 Years On, Sci. 338 (6110) (2012) 1042–1046.

[9] K. A. Dill, H. S. Chan, From Levinthal to Pathways to Funnels, Nat. Struct. Biol. 4 (1) (1997) 10–19.

[10] J. Hardy, D. J. Selkoe, The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics, Sci. 297 (5580) (2002) 353–356.

[11] M. Baba, S. Nakajo, P.-H. Tu, T. Tomita, K. Nakaya, V. Lee, J. Q. Trojanowski, T. Iwatsubo, Aggregation of Alpha-Synuclein in Lewy Bodies of Sporadic Parkinson's Disease and Dementia With Lewy Bodies, Am. J. Pathol. 152 (4) (1998) 879.

[12] D. J. Selkoe, Folding Proteins in Fatal Ways, Nature 426 (6968) (2003) 900–904.

[13] S. Takada, Gō-Ing for the Prediction of Protein Folding Mechanisms, Proc. Natl. Acad. Sci. 96 (21) (1999) 11698–11700.

[14] M. Cieplak, T. Xuan Hoang, Scaling of Folding Properties in Go Models of Proteins, J. Biol. Phys. 26 (2000) 273–294.

[15] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. De Vries, The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations, J. Phys. Chem. B 111 (27) (2007) 7812–7824.

[16] A. Davtyan, N. P. Schafer, W. Zheng, C. Clementi, P. G. Wolynes, G. A. Papoian, AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing, J. Phys. Chem. B 116 (29) (2012) 8494–8503.

[17] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al., Highly Accurate Protein Structure Prediction With AlphaFold, Nature 596 (7873) (2021) 583–589.

[18] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A. W. Nelson, A. Bridgland, et al., Improved Protein Structure Prediction Using Potentials From Deep Learning, Nature 577 (7792) (2020) 706–710.

[19] K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, et al., Highly Accurate Protein Structure Prediction for the Human Proteome, Nature 596 (7873) (2021) 590–596.

[20] L. A. Abriata, The Nobel Prize in Chemistry: Past, Present, and Future of AI in Biology, Commun. Biol. 7 (1) (2024) 1409.

[21] M. Karplus, J. A. McCammon, Molecular Dynamics Simulations of Biomolecules, Nat. Struct. Biol. 9 (9) (2002) 646–652.

[22] A. R. Leach, Molecular Modelling: Principles and Applications, Pearson education, 2001.

[23] M. Karplus, J. Kuriyan, Molecular Dynamics and Protein Function, Proc. Natl. Acad. Sci. 102 (19) (2005) 6679–6685.

[24] J. E. Jones, On the Determination of Molecular Fields.—II. From the Equation of State of a Gas, Proc. R. Soc. Lond. A 106 (738) (1924) 463–477.

[25] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, et al., CHARMM: The Biomolecular Simulation Program, J. Comput. Chem. 30 (10) (2009) 1545–1614.

[26] W. L. Jorgensen, J. Tirado-Rives, The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin, J. Am. Chem. Soc. 110 (6) (1988) 1657–1666.

[27] M. J. Dewar, E. G. Zoebisch, E. F. Healy, J. J. Stewart, Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model, J. Am. Chem. Soc. 107 (13) (1985) 3902–3909.

[28] J. J. Stewart, Optimization of Parameters for Semiempirical Methods. Iii Extension of Pm3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi, J. Comput. Chem. 12 (3) (1991) 320–341.

[29] F. Jensen, Introduction to Computational Chemistry, John wiley & sons, 2017.

[30] A. Szabo, N. S. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Courier Corporation, 1996.

[31] I. N. Levine, D. H. Busch, H. Shull, Quantum Chemistry, Vol. 6, Pearson Prentice Hall Upper Saddle River, NJ, 2009.

[32] T. Helgaker, P. Jorgensen, J. Olsen, Molecular Electronic-Structure Theory, John Wiley & Sons, 2013.

[33] J. C. Slater, Atomic Shielding Constants, Phys. Rev. 36 (1) (1930) 57.

[34] W. J. Hehre, Ab Initio Molecular Orbital Theory, Acc. Chem. Res. 9 (11) (1976) 399–406.

[35] S. F. Boys, Electronic Wave Functions-I. A General Method of Calculation for the Stationary States of Any Molecular System, Proc. R. Soc. Lond. A 200 (1063) (1950) 542–554.

[36] W. J. Hehre, R. F. Stewart, J. A. Pople, Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals, J. Chem. Phys. 51 (6) (1969) 2657–2664.

[37] J. S. Binkley, J. A. Pople, W. J. Hehre, Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements, J. Am. Chem. Soc. 102 (3) (1980) 939–947.

[38] T. Clark, J. Chandrasekhar, G. W. Spitznagel, P. V. R. Schleyer, Efficient Diffuse Function-Augmented Basis Sets for Anion Calculations. III. The 3-21+ G Basis Set for First-Row Elements, Li–F, J. Comput. Chem. 4 (3) (1983) 294–301.

[39] T. H. Dunning Jr, Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron Through Neon and Hydrogen, J. Chem. Phys. 90 (2) (1989) 1007–1023.

[40] P. A. M. Dirac, The Principles of Quantum Mechanics, Oxford university press, 1981.

[41] H. Bruus, K. Flensberg, Many-Body Quantum Theory in Condensed Matter Physics: An Introduction, Oxford university press, 2004.

[42] G. C. Wick, The Evaluation of the Collision Matrix, Phys. Rev. 80 (1950) 268–272.

[43] K. Burke, Perspective on Density Functional Theory, J. Chem. Phys. 136 (15) (2012) 150901.

[44] P. Hohenberg, W. Kohn, Inhomogeneous Electron Gas, Phys. Rev. 136 (3B) (1964) B864.

[45] W. Kohn, L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, Phys. Rev. 140 (4A) (1965) A1133.

[46] R. F. W. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, 1990.

[47] C. F. Matta, Theoretical Reconstruction Of The Electron Density Of Large Molecules From Fragments Determined As Proper Open Quantum Systems: The Properties Of The Oripavine Peo, Enkephalins, And Morphine, J. Chem. Phys. A 105 (49) (2001) 11088–11101.

[48] M. J. Timm, C. F. Matta, L. Massa, L. Huang, The Localization–Delocalization Matrix And The Electron-Density-Weighted Connectivity Matrix Of A Finite Graphene Nanoribbon Reconstructed From Kernel Fragments, J. Chem. Phys. A 118 (47) (2014) 11304–11316.

[49] C. Chang, R. W. Bader, Theoretical Construction Of A Polypeptide, J. Chem. Phys. 96 (4) (1992) 1654–1662.

[50] I. Sumar, P. W. Ayers, C. F. Matta, Electron Localization-Delocalization Matrices In The Prediction Of Pka's And Uv-Wavelengths Of Maximum Absorbance Of P-Benzoic Acids And The Definition Of Super-Atoms In Molecules, Chem. Phys. Lett. 612 (2014) 190–197.

[51] C. F. Matta, P. W. Ayers, R. Cook, Electron Localization-Delocalization Matrices, Springer, 2024.

[52] I. Sumar, R. Cook, P. W. Ayers, C. F. Matta, Aromaticity Of Rings-In-Molecules (Rims) From Electron Localization–Delocalization Matrices (Ldms), Physica Scripta 91 (1) (2016) 013001.

[53] C. M. Breneman, Transferable Atom Equivalents. Molecular Electrostatic Potentials From The Electric Multipoles Of Pro Aims Atomic Basins, in: The Application of Charge Density Research to Chemistry and Drug Design, Springer, 1991, pp. 357–358.

[54] C. M. Breneman, T. R. Thompson, M. Rhem, M. Dung, Electron Density Modeling Of Large Systems Using The Transferable Atom Equivalent Method, Computers & Chemistry 19 (3) (1995) 161–179.

[55] M. Song, C. M. Breneman, J. Bi, N. Sukumar, K. P. Bennett, S. Cramer, N. Tugcu, Prediction Of Protein Retention Times In Anion-Exchange Chromatography Systems Using Support Vector Regression, J. Chem.Inf. Comput. Sci 42 (6) (2002) 1347–1357.

[56] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, M. Uebayasi, Fragment Molecular Orbital Method: An Approximate Computational Method For Large Molecules, Chem. Phys. Lett. 313 (3-4) (1999) 701–706.

[57] D. Fedorov, K. Kitaura, The Fragment Molecular Orbital Method: Practical Applications To Large Molecular Systems, CRC press, 2009.

[58] C. Matta, L. Huang, L. Massa, Quantum Crystallography, Walter de Gruyter GmbH & Co KG, 2023.

[59] L. Massa, L. Huang, J. Karle, Quantum Crystallography And The Use Of Kernel Projector Matrices, Int. J. Quantum Chem. 56 (S29) (1995) 371–384.

[60] L. Huang, L. Massa, J. Karle, Kernel Energy Method Illustrated With Peptides, Int. J. Quantum Chem. 103 (6) (2005) 808–817.

[61] L. Huang, L. Massa, J. Karle, Kernel Energy Method: Basis Functions And Quantum Methods, Int. J. Quantum Chem. 106 (2) (2006) 447–457.

[62] L. Huang, L. Massa, C. F. Matta, A Graphene Flake Under External Electric Fields Reconstructed From Field-Perturbed Kernels, Carbon 76 (2014) 310–320.

[63] L. Huang, L. Massa, J. Karle, Quantum Kernels and Quantum Crystallography: Applications in Biochemistry, John Wiley & Sons, Ltd, 2010, Ch. 1, pp. 1–60.

[64] L. Huang, L. Massa, J. Karle, Kernel Energy Method Applied To Vesicular Stomatitis Virus Nucleoprotein, Proc. Natl. Acad. Sci. 106 (6) (2009) 1731–1736.

[65] L. Massa, T. Keith, Y. Cheng, C. F. Matta, The Kernel Energy Method Applied To Quantum Theory Of Atoms In Molecules–Energies Of Interacting Quantum Atoms, Chem. Phys. Lett. 734 (2019) 136650.

[66] V. Ganesh, R. K. Dongare, P. Balanarayan, S. R. Gadre, Molecular Tailoring Approach For Geometry Optimization Of Large Molecules: Energy Evaluation And Parallelization Strategies, J. Chem. Phys. 125 (10) (2006).

[67] F. M. Bickelhaupt, E. J. Baerends, Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry, Rev. Comput. Chem. (2000) 1–86.

[68] M. v. Hopffgarten, G. Frenking, Energy Decomposition Analysis, WIREs Comput. Mol. Sci. 2 (1) (2012) 43–62.

[69] E. D. Glendening, C. R. Landis, F. Weinhold, Natural Bond Orbital Methods, WIREs Comput. Mol. Sci. 2 (1) (2012) 1–42.

[70] R. J. Bartlett, Many-Body Perturbation Theory and Coupled Cluster Theory for Electron Correlation in Molecules, Annu. Rev. Phys. Chem. 32 (1) (1981) 359–401.

[71] T. D. Crawford, H. F. Schaefer III, An Introduction to Coupled Cluster Theory for Computational Chemists, Rev. Comput. Chem. 14 (2007) 33–136.

[72] C. Møller, M. S. Plesset, Note on an Approximation Treatment for Many-Electron Systems, Phys. Rev. 46 (7) (1934) 618.

[73] R. J. Buenker, S. D. Peyerimhoff, Individualized Configuration Selection in CI Calculations With Subsequent Energy Extrapolation, Theor. Chem. Acc. 35 (1974) 33–58.

[74] B. O. Roos, P. E. M. Siegbahn, The Direct Configuration Interaction Method From Molecular Integrals, Springer US, Boston, MA, 1977, pp. 277–318.

[75] K. Andersson, P.-Å. Malmqvist, B. O. Roos, Second-Order Perturbation Theory With a Complete Active Space Self-Consistent Field Reference Function, J. Chem. Phys. 96 (2) (1992) 1218–1226.

[76] R. J. Bartlett, M. Musiał, Coupled-Cluster Theory in Quantum Chemistry, Rev. Mod. Phys. 79 (1) (2007) 291–352.

[77] K. Raghavachari, G. W. Trucks, J. A. Pople, M. Head-Gordon, A Fifth-Order Perturbation Comparison of Electron Correlation Theories, Chem. Phys. Lett. 157 (6) (1989) 479–483.

[78] C. Riplinger, F. Neese, An Efficient and Near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method, J. Chem. Phys. 138 (3) (2013) 034106.

[79] W. Meyer, PNO–CI Studies of Electron Correlation Effects. I. Configuration Expansion by Means of Nonorthogonal Orbitals, and Application to the Ground State and Ionized States of Methane, J. Chem. Phys. 58 (3) (1973) 1017–1035.

[80] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, Phys. Rev. Lett. 69 (19) (1992) 2863.

[81] S. R. White, R. L. Martin, Ab Initio Quantum Chemistry Using the Density Matrix Renormalization Group, J. Chem. Phys. 110 (9) (1999) 4127–4130.

[82] M. A. Nielsen, I. L. Chuang, Quantum Computation and Quantum Information, Cambridge university press, 2010.

[83] U. Schollwöck, The Density-Matrix Renormalization Group in the Age of Matrix Product States, Ann. Phys. 326 (1) (2011) 96–192.

[84] S. Keller, M. Dolfi, M. Troyer, M. Reiher, An Efficient Matrix Product Operator Representation of the Quantum Chemical Hamiltonian, J. Chem. Phys. 143 (24) (2015) 244118.

[85] G. K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, S. R. White, Matrix Product Operators, Matrix Product States, and Ab Initio Density Matrix Renormalization Group Algorithms, J. Chem. Phys. 145 (1) (2016) 014102.

[86] M. Srednicki, Entropy And Area, Phys. Rev. Lett. 71 (5) (1993) 666.

[87] L. Bombelli, R. K. Koul, J. Lee, R. D. Sorkin, Quantum Source Of Entropy For Black Holes, Phys. Rev. D 34 (2) (1986) 373.

[88] J. Eisert, M. Cramer, M. B. Plenio, Colloquium: Area Laws For The Entanglement Entropy, Rev. Mod. Phys. 82 (1) (2010) 277–306.

[89] M. B. Hastings, An Area Law For One-Dimensional Quantum Systems, J. Stat. Mech.: Theory Exp. 2007 (08) (2007) P08024–P08024.

[90] W. Kohn, Density Functional And Density Matrix Method Scaling Linearly With The Number Of Atoms, Phys. Rev. Lett. 76 (17) (1996) 3168.

[91] R. F. Bader, Chemistry And The Near-Sighted Nature Of The One-Electron Density Matrix, Int. J. Quantum Chem. 56 (4) (1995) 409–419.

[92] E. Prodan, W. Kohn, Nearsightedness Of Electronic Matter, Proc. Natl. Acad. Sci. 102 (33) (2005) 11635–11638.

[93] P.-O. Löwdin, Quantum Theory Of Many-Particle Systems. I. Physical Interpretations By Means Of Density Matrices, Natural Spin-Orbitals, And Convergence Problems In The Method Of Configurational Interaction, Phys. Rev. 97 (6) (1955) 1474.

[94] P.-O. Löwdin, Quantum Theory Of Many-Particle Systems. Ii. Study Of The Ordinary Hartree-Fock Approximation, Phys. Rev. 97 (6) (1955) 1490.

[95] A. J. Coleman, Structure Of Fermion Density Matrices, Rev. Mod. Phys. 35 (3) (1963) 668.

[96] E. Davidson, Reduced Density Matrices In Quantum Chemistry, Vol. 6, Elsevier, 2012.

[97] A. J. Coleman, V. I. Yukalov, Reduced Density Matrices: Coulson's Challenge, Vol. 72, Springer Science & Business Media, 2000.

[98] J. Von Neumann, Mathematical Foundations of Quantum Mechanics: New Edition, Princeton university press, 2018.

[99] K. Boguslawski, P. Tecmer, G. Barcza, Ö. Legeza, M. Reiher, Orbital Entanglement in Bond-Formation Processes, J. Chem. Theory Comput. 9 (7) (2013) 2959–2973.

[100] L. Ding, G. Dünnweber, C. Schilling, Physical Entanglement Between Localized Orbitals, Quantum Sci. Technol. 9 (1) (2023) 015005.

[101] S. F. Boys, Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes From One Molecule to Another, Rev. Mod. Phys. 32 (2) (1960) 296.

[102] I.-M. Høyvik, B. Jansik, P. Jørgensen, Orbital Localization Using Fourth Central Moment Minimization, J. Chem. Phys. 137 (22) (2012) 224114.

[103] C. Edmiston, K. Ruedenberg, Localized Atomic and Molecular Orbitals. II, J. Chem. Phys. 43 (10) (1965) S97–S116.

[104] J. Pipek, P. G. Mezey, A Fast Intrinsic Localization Procedure Applicable for Ab Initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions, J. Chem. Phys. 90 (9) (1989) 4916–4926.

Chapter 2

Correlation in Dipeptide Systems from Quantum Information Theory Perspectives

Mostafa Javaheri Moghadam[1], Katharina Boguslawski[2], Rayna Doucet[1], Örs Legeza[3,4,5], Paweł Tecmer[2], Stijn De Baerdemacker[1,6]

[1]Department of Chemistry, University of New Brunswick, Fredericton, NB, Canada

[2]Institute of Physics, Faculty of Physics, Astronomy, and Informatics, Nicolaus Copernicus University in Toruń, Grudziądzka 5, 87-100 Toruń, Poland

[3]Strongly Correlated Systems Lendület Research Group, Wigner Research Centre for Physics, Budapest H-1525, Hungary

[4]Institute for Advanced Study, Technical University of Munich, Germany, Lichtenbergstrasse 2a, 85748 Garching, Germany

[5]Parmenides Stiftung, Hindenburgstr. 15, 82343, Pöcking, Germany

[6]Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, Canada.

Abstract

Understanding the electronic structure of amino acids is crucial for understanding protein structure and function. We interfaced quantum information indices and ab initio quantum chemistry methods to quantify the locality correlation between neighboring amino acids. Quantum chemical information from the 1-, 2-, and multi-orbital Reduced Density Matrices (RDMs) was employed to extract the mutual information (MI) and reveal local correlation patterns among constituent atoms. Notably, the MI framework is formulated in the atomic orbital (AO) representation, eliminating dependence on specific MO choices and ensuring invariance under unitary transformations, thereby providing a robust measure of electronic correlation.

2.1 Introduction

Proteins are polypeptides consisting of twenty naturally occurring amino acids that are made to perform a variety of tasks in biological systems, including signaling, substance transport, structural support, mobility, defense, etc [1]. The functionality of proteins heavily depends on their three-dimensional structure, known as tertiary structure, which itself is determined largely by the linear sequence of amino acids, known as primary structure. As described by Anfinsen's dogma, the primary structure of a protein contains the necessary information to fold into its functional three-dimensional form [2]. However, predicting the tertiary structure of proteins from microscopic principles remains a challenge to date [3]. On the one hand, proteins are sufficiently small to feel the imprint of each particular individual amino acid on its tertiary structure. On the other hand, they are also large enough such that an exact ab initio characterization from fundamental quantum mechanical principles is computationally intractable. For this reason, the protein folding problem requires physical modeling across multiple length scales.

To a large extent, popular physical models can be divided into three different length scales that coincide roughly with proteins' primary, secondary, and tertiary structures. The first and smallest level considers proteins (or parts of proteins) at the microscopic atomic and subatomic level, which requires ab initio modeling at the fundamental quantum mechanical level. At this level, the Born-Oppenheimer approximation [4] is involved in describing the molecule as a collection of nuclei that are bonded together through the quantum mechanical electron distributions [5]. The use of these models is crucial for the understanding of fine-grained interatomic interactions such as bond formation or docking processes [6, 7].

The second length scale makes abstraction of the microscopic electronic degrees of freedom and considers proteins as a collection of atoms that are held together by means of classically derivable forces. Although these models are mostly grounded in Newtonian physics, the forces that enter as input are physically informed either on a qualitative level by means of chemical intuition or on a quantitative level from *ab initio* first quantum mechanical principles [8, 9].

There has also been remarkable progress recently in machine-learning models trained on available experimental data [10, 11]. These models are mostly employed to model protein folding processes that involve global dynamics of medium-sized proteins that are beyond the capabilities of *ab initio* methods. We reach the realm of the coarse-grained models at the highest length scale. Here, several atoms are compounded together to form beads that act as individual entities. The coarse-graining of protein models becomes a necessity at the macromolecular level when the system exceeds the 100 kDa level [12, 13, 14].

At each of the three levels, it is important to incorporate correlations between the substituent degrees of freedom at the lower level. For instance, in order to accurately describe tertiary structures, it is important to understand the correlations between the constituent amino acids, for which in turn, an understanding of the electronic

properties is key [3]. The twenty naturally occurring L-amino acids are distinguished by their different side chain structures and chemical composition. Covalent bonds, hydrogen bonds, salt bridges, and disulfide bonds all contribute to the stability of the 3D structure along with hydrophobic interactions, which are pivotal in the protein's core formation [15]. Given the pivotal role of these interactions, employing a method that accurately describes the electronic structure of amino acids at a reasonable computational cost becomes imperative. Such methods provide a more realistic depiction of side-chain interactions compared to conventional force field methods, thereby enhancing our understanding of protein structure and function [16].

A variety of computational approaches have been developed to analyze molecular interactions and chemical bonding, including Energy Decomposition Analysis (EDA) [17, 18], Quantum Theory of Atoms in Molecules (QTAIM) [19], and Natural Bond Orbital (NBO) [20] methods. Each of these methods provides different insights into the nature of interatomic interactions and their contributions to molecular stability.

EDA is a valuable technique for quantitatively analyzing interaction energies between molecular fragments, making it particularly useful for comparing different molecular systems or conformations. However, EDA methods often rely on predefined fragmentations, which can introduce biases and limit their applicability. The interpretation of results can be difficult, as the division of energy components is not always clear-cut and may vary based on the chosen method and basis set. Moreover, EDA requires multiple separate calculations, one for each fragment and the entire system. This makes EDA time-consuming and computationally expensive, especially for large molecular systems [21, 22].

QTAIM theory, developed by Bader, is based on the topological analysis of electron density, which is a rigorous mathematical approach. It identifies critical points in the electron density, e.g., bond critical point, to analyze bonding and molecular structure. Matta et al. have applied QTAIM to amino acids, demonstrating how electron den-

sity topology can be used to analyze bonding patterns, hydrogen bonding interactions, and electronic delocalization in biologically relevant systems [19, 23, 24, 25, 26]. One of the main strengths of AIM is its basis-set independence. Within this framework, the transferability of atomic and group properties has also been examined, showing that apparent deviations in group energy transferability often arise from approximate treatments that violate the virial theorem, rather than from fundamental physical limitations. This highlights the importance of defining physically consistent quantities directly from the electron density [27]. In another work, Hernández-Trujillo and Bader demonstrated that molecular electron densities and associated properties can be reconstructed from transferable functional group densities defined within the QTAIM framework [28, 29].

NBO analysis provides another powerful framework for studying electronic interactions by transforming molecular wavefunctions into localized orbitals that closely resemble traditional chemical bonding concepts [30]. The advantage of NBO methods lies in their intuitive and chemically meaningful description of bonding interactions. However, NBO methods rely on predefined orbital localization schemes, which may not fully capture complex delocalization effects or non-covalent interactions. Additionally, the qualitative nature of NBO can sometimes lead to complex visual representations of bonding, particularly in large or highly delocalized systems [20].

High-performant ab initio methods have been employed to determine the interactions and correlations between amino acids and their side chains, such as CCSD-F12, CCSD(T)-F12, MP2-F12, SAPT, etc. with a relatively good level of accuracy [31, 32]. The significance of studying the quantum correlation energy of amino acids lies in its ability to provide chemical insight into the energy landscapes in which proteins exist and the mechanisms underlying their stabilization and, to some extent, their folding process. These insights arise from quantum chemical correlations between the substituent electronic degrees of freedom. These correlations can be quantified

either from an energetic, electron correlation captured by post-HF methods, or an information-theoretic theoretic, captured by the MI, perspective.

From an energetic point of view, the concept of correlation energy is defined as the difference between the exact *ab initio* ground-state energy and the ground-state energy from Hartree-Fock (HF) theory. In other words, it results from electron-electron interactions that cannot be captured by the mean field approximation. Correlation energy can be qualitatively divided into two different types: dynamic and static [33, 34, 35, 36]. Dynamic correlation relates to how electrons repel one another and can be addressed by incorporating higher-order excited-state contributions. Static correlation, on the other hand, accounts for the equivalence of configurations that are nearly degenerate with respect to the reference state. While correlation energy provides valuable insights in light of energy decomposition schemes [37], it does not allow for a deeper analysis of how different parts of the system are interconnected and how information is shared.

To address this limitation, more insightful approaches incorporate quantum information or entanglement measures [38, 39, 36, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58]. These have proven particularly insightful for the quantum mechanical characterization of chemical concepts, such as bond formation [59, 60, 45], or provide the physical foundation for powerful numerical approaches. A notable example is the Density Matrix Renormalization Group (DMRG) algorithm, of which the underlying tensor network representation is physically motivated through the notion of locality of quantum correlations [61, 62, 63, 64, 65, 66, 50].

Although the Coulomb interaction is a fundamentally non-local interaction, DMRG has found notable applications in quantum chemistry as well, mostly for multi-reference molecular systems that go significantly beyond the single-particle reference picture, e.g. in transition metal compounds [39, 67, 68, 69, 39, 70]. Here, quantum information theory can be efficiently employed as a post-analysis tool to elucidate the

electronic structure and correlations in molecular systems [36, 44, 46]. For instance, the MI index is a measure of the quantum correlations between any two individual subsystems within the molecule. [38] A natural and popular choice of subsystem is the single-electron spin-orbital, given its foundational role within chemistry. Indeed, the MI between individual molecular or natural spin-orbitals provides a good measure for the multi-reference character of the molecular system.

In one study it was shown that the nature of correlation energy, whether it is dynamic or static, can be reflected in the entanglement patterns of orbitals [36].

In another work, some of us assessed the amount of entanglement between molecular orbitals (MOs) in diatomic systems to elucidate bond formation and dissociation processes [59, 60, 45, 49, 47].

Furthermore, MI indices are also informative as optimization tools for practical electronic structure computations with increasing levels of sophistication. For instance, they can significantly help simplify the underlying tensor network wavefunction in DMRG by optimizing the ordering of the spin-orbitals such that quantum correlations are strongest in adjacent orbitals [66, 71, 72, 73, 74, 75]. However, through the optimization and reordering of MOs, a large portion of the intuitive spatial correlations within a molecule get obfuscated in the process, leading to a loss of the notion of locality. As such, one can argue that DMRG has not been applied to its potential in molecular systems: 1D extensive molecules with local quantum correlations. Examples of such systems that come naturally to mind are polymers and polypeptides or proteins.

With this paper, we embark on a research project to capture the electronic structure of extensive effective 1D molecular systems within a tensor network formalism. Therefore, two major developments need to take place. First, the establishment of the local character of quantum correlations, derived from MI, in polypeptides and polymers, and second the identification of the optimal tensor network architecture.

In the present paper, we focus on the former aspect, evaluating and confirming the locality of quantum correlations in peptide systems using MI indices. More specifically, we compute the MI between local AOs, rather than canonical MOs or localized orbitals (obtained by localizing MOs), as was done in previous works (see, for instance, [76, 77, 51, 52, 57]), and compound them to fully interatomic MI indices. The choice of AOs over localized molecular orbitals (LMOs) enables a systematic, atom-centered analysis of quantum correlations, derived from MI. LMOs introduce arbitrariness in localization methods (e.g., choice of the Foster-Boys [78] or Pipek-Mezey [79] method), potentially emphasizing bonding regions while downplaying weaker interactions.

A potential concern is that orthogonalization tails in AOs may affect MI; however, these effects arise systematically from basis set construction. Using correlation-consistent basis sets (cc-pVXZ) and increasing the resolution of the basis in a grid, we ensured consistency across systems. The observed MI trends, higher for double bonds than single bonds, align with chemical expectations, suggesting minimal artificial influence.

Our results confirm the locality of quantum correlations, derived from MI, in these systems. This is in line with recent work by Pekker et. al. which showed that the matrix elements of 1- and 2-RDMs decrease rapidly as a function of the distance and changes in local conformation only impact density matrices locally [80].

Furthermore, our findings align with the observation that LMOs can effectively capture the essential electronic structure of these systems, reinforcing the idea that quantum correlations remain predominantly local despite extended conjugation effects [81, 82]. Additionally, the presence of long-distance entanglement in molecular systems, particularly at low and high electron densities, suggests that the locality of quantum correlations, derived from MI, can be influenced by electronic structure and interorbital interactions, as observed in tight-binding and hydrogen ring models [57].

Computations were performed at the DMRG, HF, and Density Functional Theory

(DFT) levels, from which density matrices are obtained from the Matrix Product State (MPS), HF and Kohn-Sham Slater determinant respectively.

Interestingly, we observe a large degree of correlation and consistency between the results of these three vastly distinct methods. This consistency across computational approaches highlights that the observed locality of quantum correlations is a fundamental property of electronic structure rather than an artifact of a specific method, further validating the robustness of our findings.

The paper is organized as follows. In the Methods section (Section 2.2), we review the concept and necessary ingredients of MI in quantum chemistry and provide computational details on the test pool of peptides. The use of MI as a locality measure is then discussed in the proceeding Results section (Section 2.3). We present our conclusions in the final section (Section 2.4.)

2.2 Methods

2.2.1 Mutual information

Our approach is largely adopted from [38, 39, 59], differing only in that we evaluate orbital entanglement and correlation between AOs rather than MOs. Specifically, our methodology includes an assessment of the correlation between any two AOs and the entanglement of a single orbital with every other orbital. Entanglement between two subsystems of a bi-partite system being in a pure state

$$|\Psi\rangle = \sum_{n,e} \Psi_{n,e} |n\rangle \otimes |e\rangle, \tag{2.1}$$

can be quantified by the von Neumann entropy, obtained from the RDM of the subsystem. Here, n represents a chosen AO (the subsystem), while e corresponds to the remaining orbitals in the system (the environment). The RDM of the subsystem is

given by

$$\rho = \mathrm{Tr}_{|e\rangle}|\Psi\rangle\langle\Psi| \tag{2.2}$$

in which ρ is a Hermitian positive semidefinite operator with the dimension of the Hilbert space of the subspace. All eigenvalues ω_α of ρ are positive so the entanglement entropy is well defined

$$s_n = -\sum_\alpha \omega_\alpha \ln\ \omega_\alpha. \tag{2.3}$$

All entropies will be measured in natural units (nat). In our analysis of entanglement three entropic and correlation measures are investigated: (i) The single-orbital entropy [66],

$$s(1)_i = -\sum_{\alpha=1}^{4} w_{\alpha,i} \ln w_{\alpha,i}, \tag{2.4}$$

which quantifies the amount of entanglement between a specific spin-orbital i and the other orbitals (Figure 2.1a). Here, the one-orbital RDM ρ_i is a four-dimensional density matrix with $\mathcal{H}_i = \mathrm{span}\{\ |-\rangle, |\uparrow\rangle, |\downarrow\rangle, |\updownarrow\rangle\ \}$, that can be obtained by tracing out all other orbital contributions from the many particles RDM except those of orbital i. The diagonal matrix elements of the one-orbital RDM for (spatial) orbital i can be calculated from the expectation values of generalized correlation functions [38, 48] which can also be expressed via proper combinations of the expectation value of standard occupation number operators [59].

The second (ii) entropy measure $s(2)_{ij}$ represents the entropy of two orbitals i and j, and can be determined by

$$s(2)_{ij} = -\sum_{\alpha=1}^{16} w_{\alpha,ij} \ln w_{\alpha,ij}. \tag{2.5}$$

The $w_{\alpha,ij}$ are the eigenvalues of the two-orbital RDM ρ_{ij} and can be obtained by tracing out all other orbitals contribution from the many-particle density matrix except orbitals i and j. As the pair of spatial orbitals $\{i, j\}$ can be occupied by zero to four

electrons, the $\rho_{i,j}$ has a dimension of 16, $\mathcal{H}_{i,j} = \{|-,-\rangle, |-,\uparrow\rangle, |-,\downarrow\rangle, \cdots, |\updownarrow,\updownarrow\rangle\}$. The non-zero elements of ρ_{ij} can again be calculated from generalized correlation functions leading to a block sparse structure. The underlying algebra can also be expressed in terms of standard particle creation and annihilation operators. For more details we guide readers to Refs. [38, 59, 48, 55]

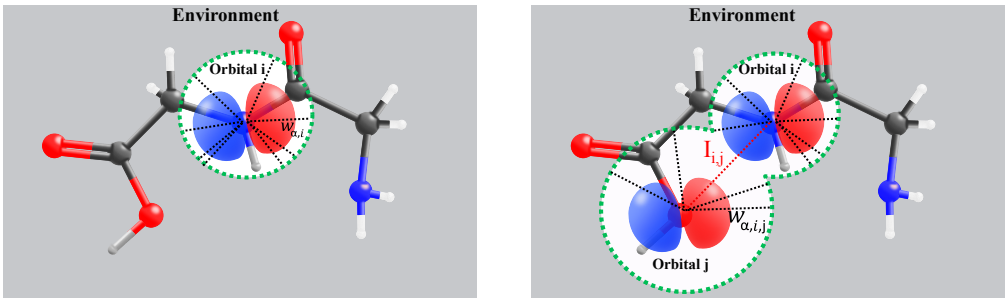
Third, the MI between two orbitals (iii) [38]

$$I_{ij} = s(2)_{ij} - s(1)_i - s(1)_j, \tag{2.6}$$

quantifies the degree of correlation (that has both quantum and classical origins) between the orbitals i and j as they are embedded in the whole system. In the conventional MO picture, both single-orbital entropy ($s(1)$) and two-orbital entropy ($s(2)$) are zero for a single Slater determinant wavefunction, as each MO is either fully occupied or unoccupied, with no intrinsic uncertainty or correlation in the post-HF sense. In multi-determinant wavefunctions, these entropies become nonzero; we refer the readers to Refs [44, 83, 84] for further details. By working in the AO picture, however, these entropy measures remain meaningful even for single Slater determinants, allowing us to quantify orbital correlations within HF or DFT. In this framework single-orbital entropy ($s(1)$) reflects how much an AO is entangled to the AO bath, whereas $s(2)$ quantifies how much an orbital pair is entangled with the orbital bath. The MI (I_{ij}) quantifies the correlation between two AOs (both classical and quantum).This AO-based approach enables the assessment of MI and entropy values even in HF and DFT, providing a valuable (information theoretical) perspective on electronic correlations where traditional electron correlation measures (such as in post-HF approaches) are absent.

Here, it is noteworthy to emphasize the key distinction between $s(2)_{ij}$ and I_{ij}. $s(2)_{ij}$ computes the entanglement between the orbital pair i and j with all other active-space

orbitals (the environment) while I_{ij} quantifies only the specific correlation between orbitals i and j, within the environment. This allows for a more targeted analysis of the direct correlation between these two orbitals, see Figure 2.1b. It is good to keep in mind that the definition of MI can be conceptually extended beyond the single spin-orbital level. For instance, we will discuss the MI between blocks of several spin-orbitals, typically associated with atoms, which we will refer to as $I^{(\mathrm{b})}_{AB}$, or the MI between block-A and block-B.



(a) (b)

Figure 2.1: Schematic graphical representation of 1- and 2-orbitals RDMs in the Glycine-Glycine (GlyGly) dipeptide system. The MI, I_{ij}, between i and j is the difference between the sum of the entropies of the individual orbitals i, j and the pair orbital entropy.

The computation of the MI index I_{ij} involves the expectation value of a variety of n-body operators $\mathcal{O}^{(k)}_i$ listed in Refs. [38, 59, 48, 55] from which the orbital-RDMs can be determined. These orbital-RDMs can be further resolved using up to 4-particle RDMs [44, 83]. Whereas high-quality RDMs or, eventually, orbital-RDMs can be constructed from sophisticated electronic structure theories, such as DMRG, this is not practical when going towards extensive systems. For this reason, we will investigate the effectiveness of lower-quality wavefunctions based on single-particle assumptions, i.e., mean-field solutions, like HF theory or Kohn-Sham Density Functional Theory (KS-DFT), and validate the results against expensive DMRG results for smaller peptide systems. The benefit of employing Slater determinant wavefunctions is the existence and applicability of Wick's theorem [85], significantly reducing the computation

cost of n-body operator expectation values. An example of Wick's theorem at work is provided in Appendix 2.6.1. The obvious price to pay for this simplification is that it does not apply to multi-reference systems with strong quantum correlations, such as transition metal compounds. However, our ultimate focus is to verify the locality of quantum correlations in extensive organic systems, for which self-consistent field methods have proven to provide sufficiently accurate results for different purposes [86, 87, 88].

An important aspect of the MI indices is that the theory of quantum entanglement and correlation measures is built on the premises of mutually orthogonal Hilbert spaces. This is not satisfied by construction when working with traditional AOs, because of the non-zero overlaps between different AOs. For our purposes, it is essential to work with orbitals that are both localized and orthogonal. Many localization schemes exist in the literature for organic chemistry purposes [81], however, we will work here with the straightforward so-called symmetric orthogonalization, defined by [89]

$$c^{\dagger}_{j\sigma} = \sum_{i=1}^{L} S_{ij}^{-\frac{1}{2}} a^{\dagger}_{i\sigma} \tag{2.7}$$

in which S is the overlap matrix of the non-orthogonal AO basis, as they are sufficiently close to the original AO basis. Other localization procedures can be explored as well, as long as they fulfill the condition of orthonormality. MI indices between AOs will provide a measure for locality of quantum correlations, however, it is overly fine-grained from a molecular perspective, for which full atoms are better suited as building blocks. Therefore, it is important to extend the MI index I_{ij} from the orbital to the atomic picture. For this, one can define the full set of AOs associated with one particular atom as the subsystem, or block, of interest. Obviously, this is strongly dependent on the choice of basis set and can become prohibitively expensive for a large basis set, as the RDM of the 2-atom subsystem requires a computation that

scales as a diatomic full Configuration Interaction for each individual charge- or particle number sector. To overcome this limitation we employed an intuitively effective approximation, which involves approximating the existing MI between atoms I_{AB} by summing over all MI that exists between the constituent orbitals of those atoms. This can be expressed as

$$I_{AB} \approx I_{AB}^{(a)} = \sum_{i=1}^{L_A} \sum_{j=1}^{L_B} I_{ij}, \tag{2.8}$$

where L_A and L_B represent the total number of spatial orbitals in atoms A and B, respectively, and I_{ij} is the MI between orbital i and j. A graphical depiction of this approximation is given in Figure 2.2, where the total amount of information between atom A and B is represented by summing over all orbital-wise information that exists between these two atoms (blue lines in Figure 2.2). From now on we refer to the $I_{AB}^{(a)}$ as the (approximate) AMI which was used to measure the amount of correlation between constituent atoms of molecules. The validity of the approximation is investigated in Section 2.3.

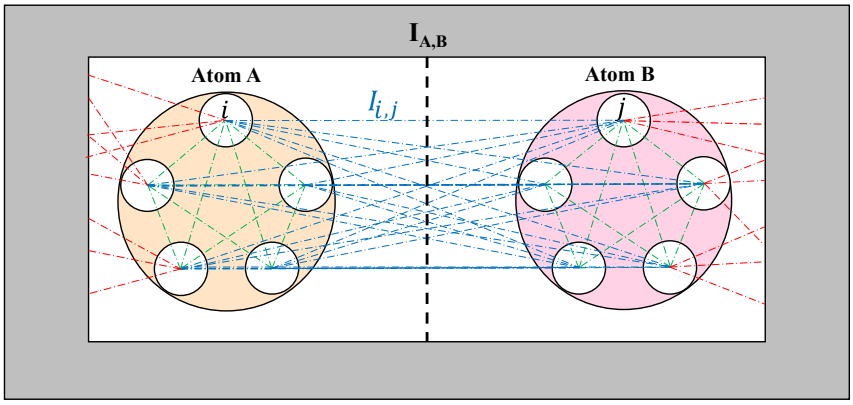


Figure 2.2: Cartoon representation of the summation wise $I_{AB}^{(a)}$ between atom A and B as a sum over all MI that exists between the spatial orbitals of atom A and of atom B. Each A and B atom consists of 5 spatial orbitals. The thickness of the lines shows the magnitude of MI. Blue lines show the MI between atoms A and B, green lines indicate the MI between spatial orbitals within the atoms, and red lines indicate the correlation between the atoms and the remaining system.

The practical utility of this approach is demonstrated on the set of all 400 possible

dipeptides consisting of two of the twenty natural amino acids. This choice of test set is motivated by two reasons. First, dipeptides are regarded as the smallest polypeptide chains, allowing for a more manageable computational load. Second, the use of all possible combinations of the twenty naturally occurring amino acids allowed for the creation of a large and sufficiently diverse library of 400 dipeptides. The results of our work not only allow for the validation of locality in organic peptide systems but also shed light on the bonding patterns and the strength of connections between atoms in dipeptides, offering a blueprint to extend such connections to larger polypeptides and even proteins.

To put our work in a broader perspective, traditional approaches in quantum chemistry have long relied on qualitative representations of MOs [19]. These representations are valuable for conceptualizing the qualitative concept of bonding and antibonding orbitals, as well as the general distribution of electrons within a molecule. In contrast, the MI provides a quantitative measure that indicates the degree of correlation between orbitals, offering a more detailed understanding of the electronic interactions within molecules. Expressed in natural logarithmic units (nat), MI quantifies the correlation strength between orbital pairs. Previous studies have leveraged orbital correlation measures to analyze electron correlation effects, chemical reactions, and bonding patterns [44, 39, 36, 59, 43, 41, 45, 49, 47, 56, 90, 82, 58]. While these studies are valuable for understanding strong correlations, derived from MI, they do not provide an atom-centric perspective. In contrast, the analysis becomes more spatially resolved by focusing on AOs, providing insight into how correlation is distributed spatially throughout extensive systems. This can be particularly beneficial when investigating local bonding patterns or the influence of specific atoms on the overall electronic structure. Moreover, MI calculated between AOs can be more useful where one is interested in the strength and nature of correlations between individual atoms rather than a global assessment of the entire molecule. This finer resolution

can be crucial when studying complex molecular systems where electron interactions are highly localized.

2.2.2 Dipeptides preparation and computational details

We curated a comprehensive set of electronic structure calculations on 400 dipeptides at different levels of theory. The neutral termini were selected for each dipeptide and they were subjected to initial geometry optimization using semi-empirical methods, specifically AM1 and PM3 methods. Subsequently, a deeper level of geometry optimization was applied utilizing the hybrid density functional B3LYP [91] with the 6-31G(d) basis set. To study the dipeptides' $I_{AB}^{(a)}$ the range-separated hybrid density functional ωB97M-V [92] was employed, chosen for its reliability in capturing diverse interaction types based on a comprehensive study of 200 different functionals by Mardirossian et al [93]. All DMRG calculations, including the calculation of the MI and single orbital entropies, have been performed with the Budapest DMRG program [94]. The 6-31G(d) basis set was used and the full active space calculations, including all spatial orbitals, were employed. All DMRG calculations have been converged until chemical accuracy ($\Delta E < 1.5$ mHa) has been reached within consecutive sweeps. We should stress that we are primarily interested in generating reference MI distributions, which typically converge much faster than the corresponding (total) energies. Though the total DMRG energies might not be fully converged, the corresponding MI patterns are not expected to change significantly. See also Appendix 2.6.2 for a comparison of the MI patterns for a different number of renormalized states. We employed the dynamic block state selection (DBSS) protocol [66] to dynamically choose the number of renormalized active-system states D according to a predefined threshold value for the quantum information loss. Specifically, the minimum value for D was set to 1024, the maximum value was set to 2048, while the threshold for the quantum information loss was set to 10^{-5} in all DBSS-based DMRG calculations. All DFT calculations

were performed by the Orca program package version 5.0.4 [95] on Digital Research Alliance of Canada.

2.3 Results and discussions

Visualizations are a major asset of MI indices, facilitating intuition and interpretation of chemical structure at a glance. The majority of previous investigations into MI [96, 39, 59, 97], represented MI in circular plots, allowing for the elucidation of complex cases of strong correlation, derived from MI, in chemical systems. In the present study, however, the use of localized AOs instead of MOs facilitates quick visual inspection of the entanglement and correlation patterns between constituent atoms, in similar ways as presented earlier between individual orbitals [51] or lattice sites [98].

Figure 2.3 compares the MI, plotted in a circle, versus $I_{AB}^{(a)}$, plotted based on the atoms' spatial position, calculated by a DBSS-based DMRG algorithm [66] on Glycine. Similar to before [59], we use the thickness of the lines to visualize the $I_{AB}^{(a)}$ between atoms. In these SPatial Atom-Wise Network (SPAWN) plots, the thick lines are strongly correlated atoms whereas the thin lines denote weakly correlated atoms. Already on a high qualitative level, we note that covalently bonded atoms have high $I_{AB}^{(a)}$ with a further visual classification between double and single bonds, differentiating for instance between carbonyl and hydroxy bonds. This allows for a quick visual inspection of chemically significant structures.

It should be kept in mind at all times that the thickness of the lines between the atoms in the SPAWN plots are dictated by the $I_{AB}^{(a)}$, and are not informed by any prior chemical input about bonding structures, etc. from the user, as in standard chemical drawings. To enhance clarity, the layout prioritizes the representation of AMI values and correlations, occasionally at the expense of geometric accuracy.

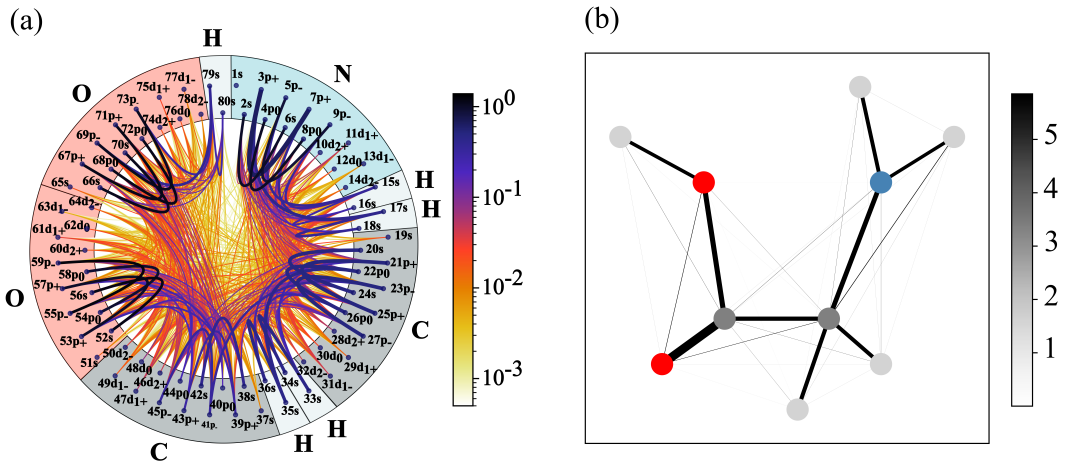


Figure 2.3: Circular (a) and spatial atom-wise (b) representations of (a) MI and (b) $I^{(a)}_{AB}$ between AOs in Glycine amino acids obtained from a DBSS-based DMRG calculation. Each AO in (a) is labeled according to its associated atom, with nitrogen, carbon, and oxygen possessing 14 AOs each, while hydrogen atoms have 2 AOs in the 6-31G(d) basis set. The line thickness and color gradient indicate the degree of correlation. The circular graph (a) was generated by the Orbital Entanglement module of the PyBEST software package [99, 100] and then labeled accordingly to reflect the atom-wise composition of the MI. AOs are combined and grouped under their respective atoms in (b), with oxygen represented by red nodes, carbon by dark gray, nitrogen by blue, and hydrogen by light gray. The thickness and shade of the lines reflect the degree of MI.

By implementing this approach we explore the entanglement in 400 dipeptides and visualize the atomic correlations using SPAWN plots. Since displaying all 400 dipeptides would be overwhelming, we present a diverse and meaningful selection of twelve SPAWN graphs in Figure 2.4, chosen to capture key trends across all 20 amino acids. While BA and AB dipeptides share the same amino acid composition, their electronic structure, and quantum correlations, derived from MI, differ due to the termini, where one end has a carboxyl group and the other an amino group, influencing their bonding environment.

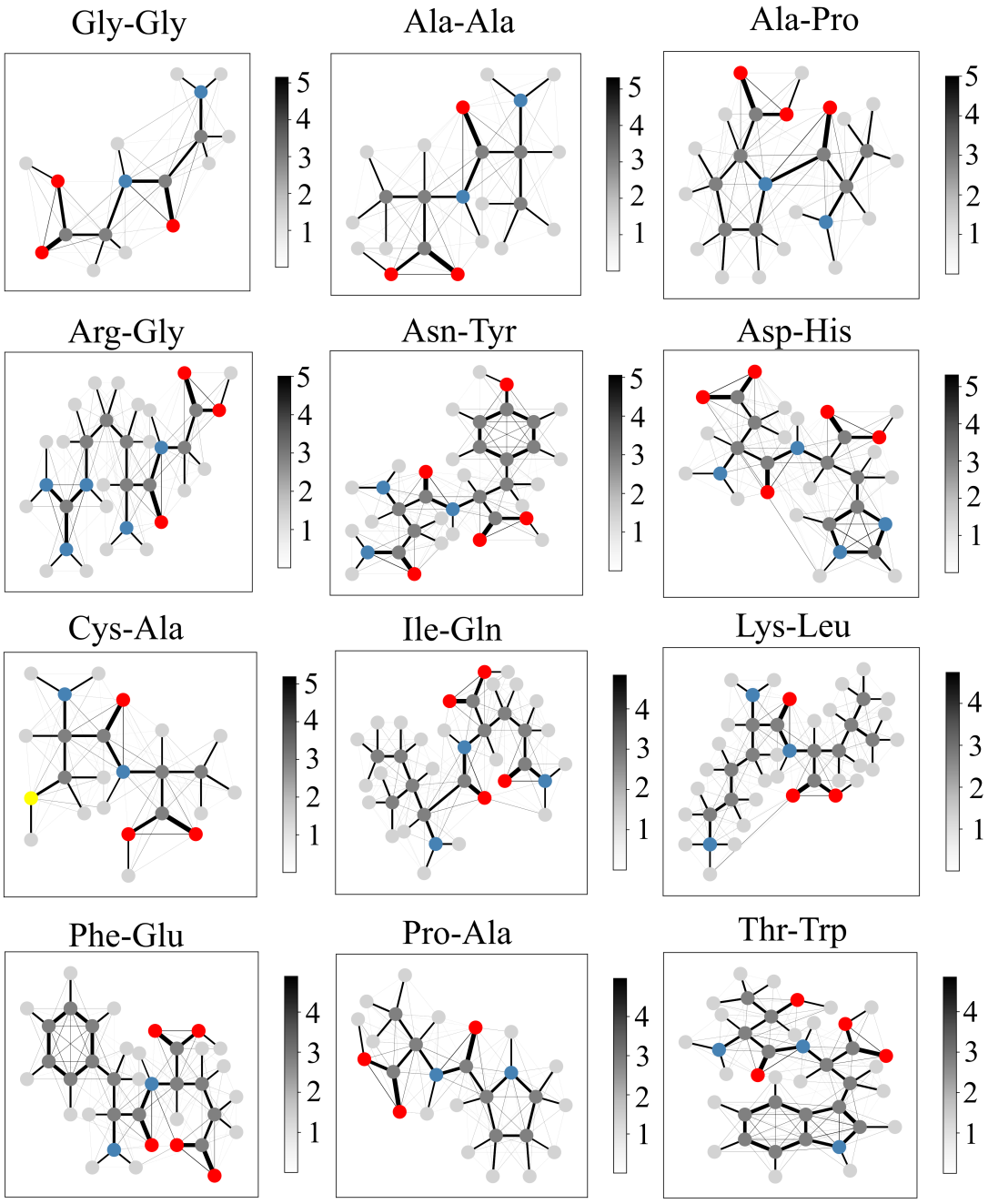


Figure 2.4: SPAWN plots for 12 selected dipeptides. Geometries were optimized at the B3LYP level, and then ωB97M-V was used to compute AMI. Oxygen is represented by red nodes, carbon by dark gray, nitrogen by blue, sulfur by yellow, and hydrogen by light gray. The thickness and color gradient of the lines reflect the degree of correlation.

A complete dataset including SPAWN plots for the 400 dipeptides can be found at https://github.com/QuNB-Repo/Dipeptides. A quick visual inspection of the

twelve SPAWN plots shows that $I_{AB}^{(a)}$ is generally consistent with chemical rules involving covalent and non-covalent interatomic correlations. To ensure clarity in the presentation, all subsequent computational results are obtained at the DFT level unless explicitly stated otherwise.

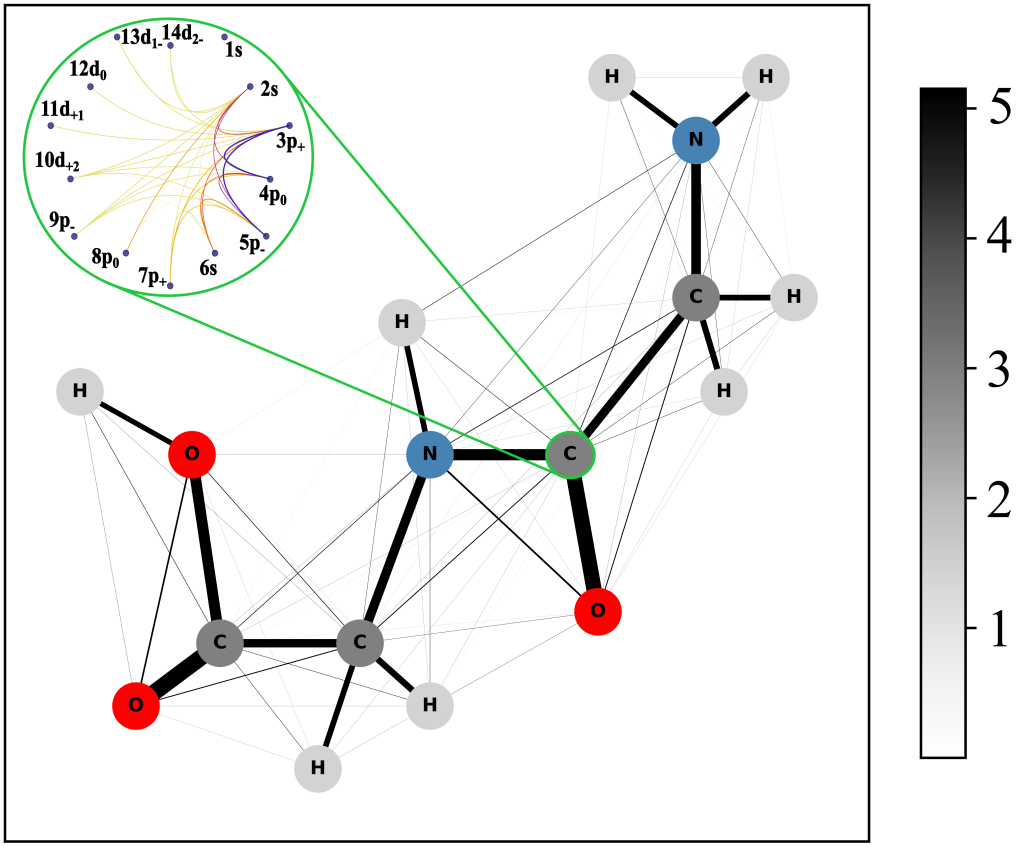


Figure 2.5: SPAWN plot for the Glycine-Glycine dipeptide. Carbon, Nitrogen, Oxygen and Hydrogen atoms are represented by dark gray, blue, red and light gray dots respectively. The connecting lines show the degree of correlation among atoms. The distance of the peptide bond is optimized to be 1.36 Å. The MI of 14 AOs is illustrated only for one carbon atom to avoid visual cluttering. The thickness and color gradient of the lines reflects the degree of AMI$\in [10^{-3}, 5]$.

We break down the analysis for the particular case of the Glycine-Glycine dipeptide (top left dipeptide in Figure 2.4) in Figures 2.5 and 2.6. Figure 2.5 shows the $I_{AB}^{(a)}$ obtained from a DFT calculation for 70 electrons in full active space of 142 spatial orbitals. The connections between atoms represent $I_{AB}^{(a)}$ indices and the thickness of the lines depicts the magnitude of the correlation between connected atoms. To avoid

visual cluttering, the intra-atomic MIs are illustrated only for one atom.

The stretch of the dipeptide bond in Gly-Gly provides a first test to validate the locality of the $I_{AB}^{(a)}$ index. The $I_{AB}^{(a)}$ between the nitrogen and carbon atom in the peptide bond is plotted in Figure 2.6c as the distance between the atoms is increased while maintaining the geometries of the individual amino acids (see Figure 2.6a).

The computed AMI between the atoms forming the peptide bond in Gly-Gly was 3.52 nat at an equilibrium distance of 1.36 Å. As the bond length increases, the $I_{AB}^{(a)}$ between the nitrogen and carbon atoms in the peptide bond decreases, reflecting weakened electronic correlation between these two atoms (see blue circles in Figure 2.6c). However, the rate of this reduction slowed down at 2.4 Å, where the total energy of the system, shown in Figure 2.6d, indicates the presence of a transition into a dissociated state of two unsaturated radicals.

Notably, we found that AMI remains nonzero for radical fragments even at large bond distances, suggesting a persistent dependence on electronic structure correlation, regardless of the distance between the atoms. In contrast, when the peptide bond is indeed saturated with a water molecule (see Figure 2.6b) the rate of $I_{AB}^{(a)}$ reduction increases significantly around 2.0 Å, indicating the onset of rapid electronic destabilization. Beyond 2.4 Å, as the bond length stretches further, the separated saturated amino acids exhibit weak van der Waals interactions, which are reflected in small but non-zero $I_{AB}^{(a)}$ values ranging between 10^{-1} nat and 10^{-5} nat.

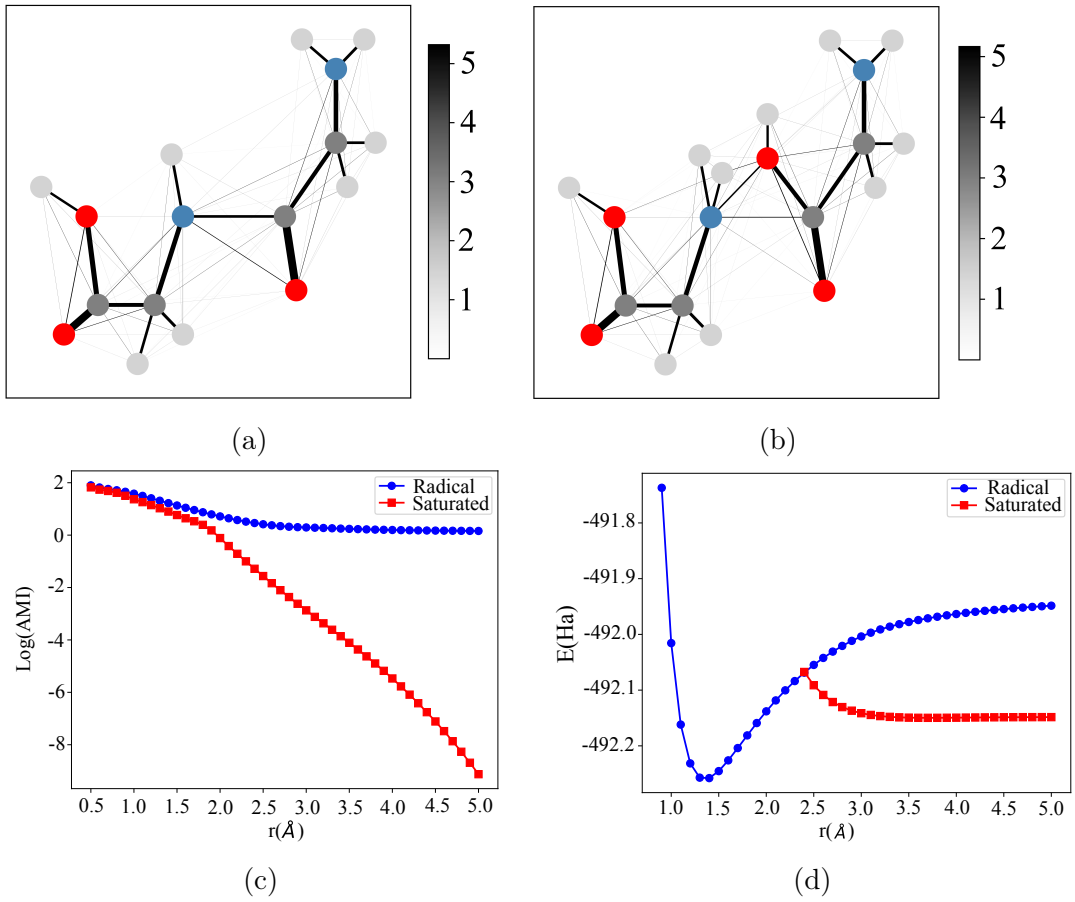


Figure 2.6: SPAWN plots for a stretched Gly-Gly dipeptide system with (a) unsaturated peptide bond at 2.2 Å and (b) saturated peptide bond with a water molecule. The changes in $I_{AB}^{(a)}$ (AMI) between the nitrogen and carbon atoms in the peptide bond (c) and total energy of the system (d) as a function of intermolecular distance.

The reduction in $I_{AB}^{(a)}$ observed in a stretched peptide bond highlights the locality of electron correlation. This phenomenon is further corroborated by analyzing $I_{AB}^{(a)}$ based on the distances between various pairs of atoms in dipeptides. Figure 2.7 displays the $\log(I_{AB}^{(a)})$ between all possible atom pairs for the same twelve selected dipeptides as Figure 2.4.

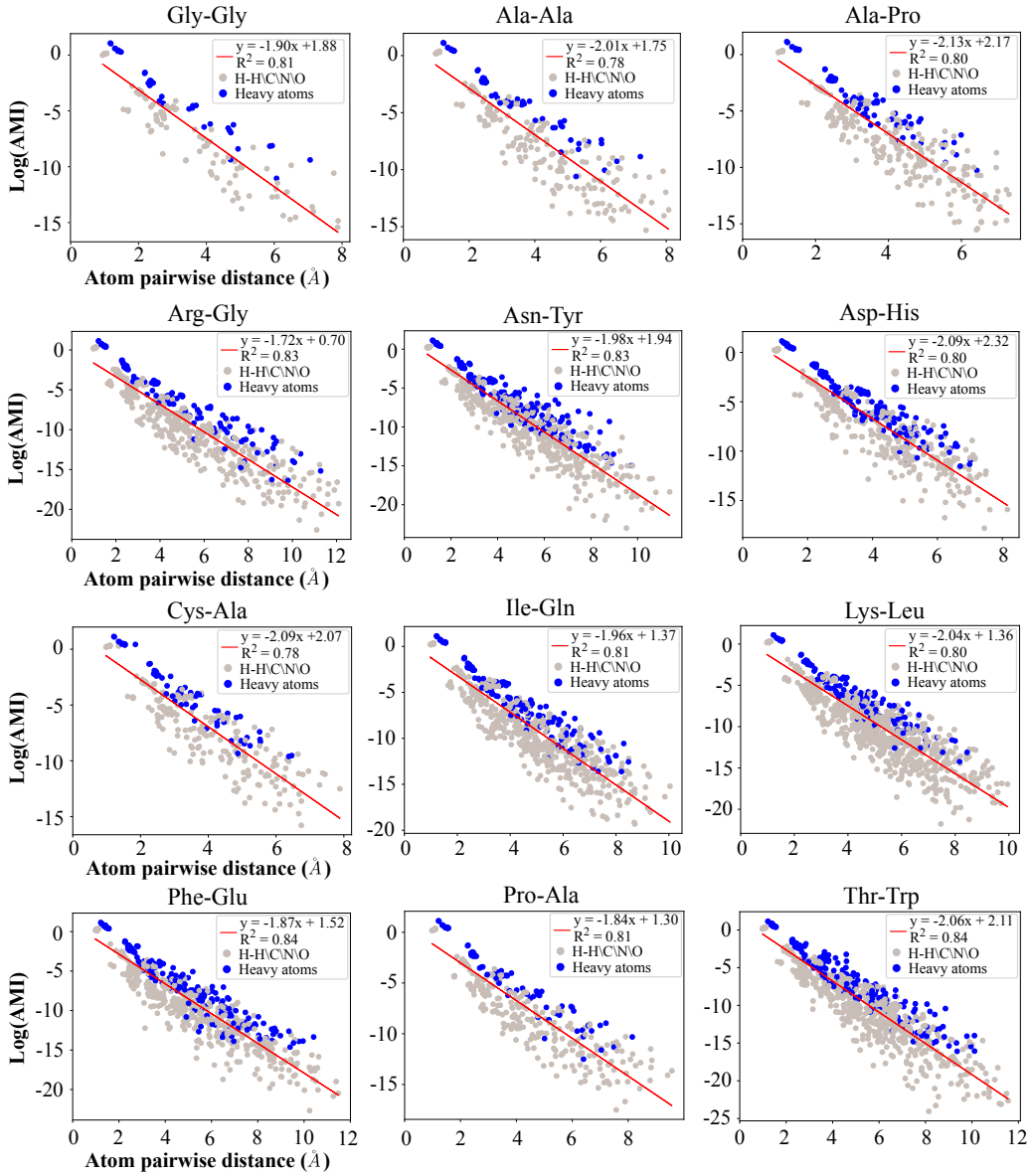


Figure 2.7: Pairwise $I_{AB}^{(a)}$ between all atoms of 12 different dipeptides derived from ωB97M-V after geometry optimization using B3LYP. Blue dots represent pairwise atomic $I_{AB}^{(a)}$ (AMI) when both atoms are heavy (carbon, nitrogen, oxygen, and sulfur), while gray dots represent pairwise atomic $I_{AB}^{(a)}$ involving at least one hydrogen atom.

From this Figure, the correlation between $I_{AB}^{(a)}$ and interatomic distance is clear, which can be understood from a chemical point of view. At the top left corner of each of the twelve figures, we find the atom pairs at short interatomic (bond) distances with high $I_{AB}^{(a)}$, which are easily identified as the covalently bonded atom pairs. At larger dis-

tances, we find atom pairs that interact predominantly via non-covalent interactions, which are characterized by a decreased level of correlation, derived from MI. More importantly, we observe a consistent linear correlation in the semi-log plot, pointing towards an exponential decay of $I_{AB}^{(a)}$ or quantum correlations as a function of distance, underscoring the nearsightedness or locality of interatomic interactions. The R^2 linear regression score provides a measure of the validity of the exponential decay, which is around $R^2 \sim 0.80$ for all presented systems, providing a reasonable confirmation of the exponential decay relation. This exponential decay behavior closely resembles trends observed in QTAIM delocalization indices and localization-delocalization matrices (LDMs), which measure electron sharing between atomic basins. However, in our approach correlation network among AOs emerges directly from the wavefunction without relying on predefined atomic basins, offering a more flexible and generalizable framework [101, 102].

An important aspect to consider is whether the exponential relation is unique to dipeptide systems or would fail to emerge for more extensive systems. For this reason, the Alanine tetra-peptide (Ala-Ala-Ala-Ala) was studied as well, see Figure 2.8. The choice for this particular tetra-peptide is deliberate as the Alanine amino acid, with its small methyl side chain, has a compact geometry where the backbone length and side chain width are comparable, making it less polymeric compared to bulkier amino acids like Lysine. However, in a tetra-peptide, the system becomes more elongated, better resembling an extended polypeptide chain. We retrieve a similar R^2 score for the tetra-peptide as the dipeptide, pointing out that $I_{AB}^{(a)}$ indices should be safely transferable to larger systems. These patterns confirm that quantum correlations, derived from MI, weaken exponentially as atomic distance increases. These findings are crucial for understanding the behavior of larger polypeptides, demonstrating that electronic interactions weaken predictably with increasing atomic distances, thereby reinforcing the concept of electronic nearsightedness in complex molecular structures.

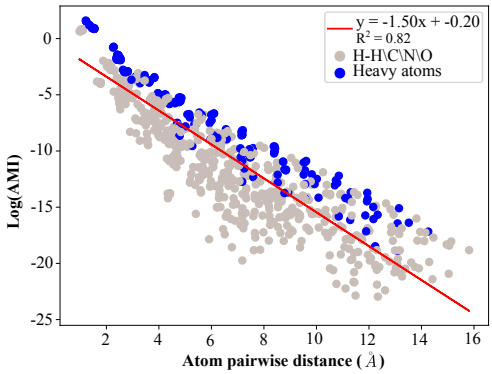


Figure 2.8: The atomic pairwise $I_{AB}^{(a)}$ (AMI), derived from ωB97M-V, as a function of distance in Alanine tetra-peptide. Blue dots represent pairwise atomic $I_{AB}^{(a)}$ when both atoms are heavy (carbon, nitrogen, and oxygen), while gray dots represent pairwise atomic $I_{AB}^{(a)}$ involving at least one hydrogen atom.

At this particular point, it is fair to note that all computational results have been obtained at the DFT level of theory. The choice of a computationally inexpensive method has been deliberate for twofold reasons. First, DFT remains the most popular *ab initio* approach for biomolecular systems, regardless of CCSD(T) recently emerging as a valid competitor for extensive bio-molecular systems [103, 104, 105, 106]. Second, despite the computational scaling, it allowed for the generation of the complete dataset of dipeptide systems that is large enough to capture statistical trends. However, it is crucial to dissect some of the conclusions at a higher level of theory. Because of the readily available reliable RDMs within its formalism, the choice fell on DMRG. Two specific dipeptides, Gly-Gly and Ala-Gly, were selected for comparison with the more precise and computationally demanding DMRG calculations. These dipeptides were chosen because they are the smallest, making it feasible to perform DMRG calculations using a full Complete Active Space (CAS) approach based on AOs. We have illustrated only the Gly-Gly result, as the result is closely comparable to that of the Ala-Gly dipeptide, however, the Ala-Gly result is presented in Figure 2.14 in Appendix 2.6.2 for the interested reader.

Figure 2.9 compares the SPAWN graphs of Gly-Gly dipeptides obtained from DMRG

and DFT calculations, including ωB97M-V, B3LYP, and HF methods. Despite the lower-quality wavefunctions in DFT and HF compared to DMRG, the $I_{AB}^{(a)}$ patterns from DMRG show remarkable similarity to those from DFT and HF.

Figure 2.10 provides a more quantitative comparison of the $I_{AB}^{(a)}$ indices derived from DMRG with those from HF, B3LYP, and ωB97M-V for the Gly-Gly dipeptide. The plots reveal a high degree of alignment between the lower-quality wave functions (DFT and HF) and the high-quality DMRG results.

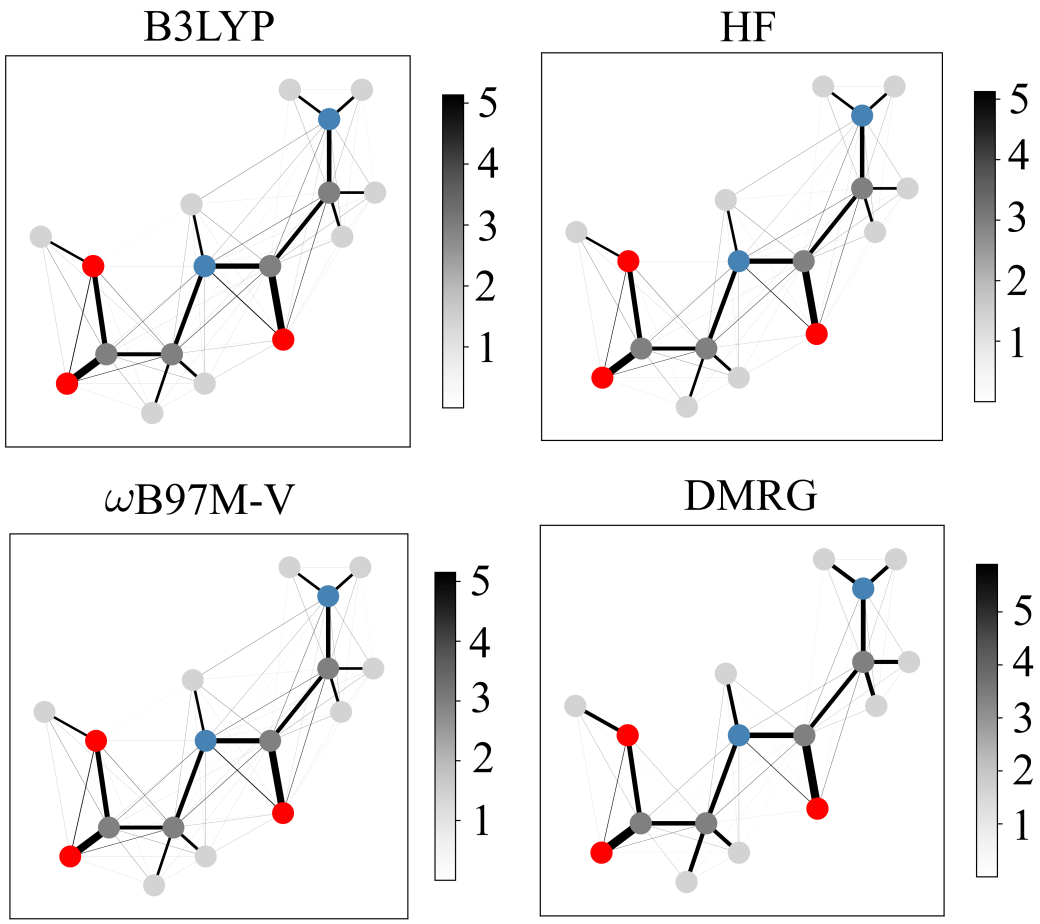


Figure 2.9: SPAWN plots for Gly-Gly dipeptide at the B3LYP, HF, ωB97M-V, and DMRG($D = 2048$) level of theory.

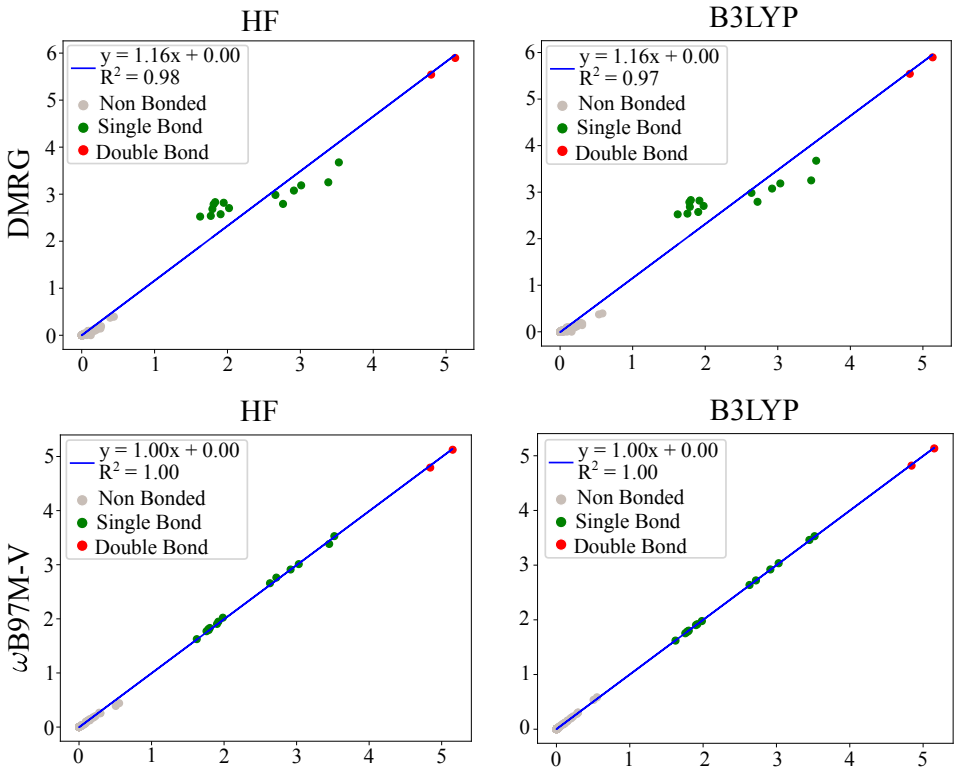


Figure 2.10: Comparison of AMI $I_{AB}^{(a)}$ for the Gly-Gly dipeptide at different levels of theory. Upper panels compare DMRG against HF and B3LYP, and lower panels confront ωB97M-V against HF and B3LYP (lower panel). Double and single bonds are denoted by red (upper right) and green dots (center middle), respectively, whereas non-covalently bonded atom pairs (lower right) are pictured with gray dots.

We observe a striking correlation between the self-consistent field approaches and the DMRG, in which the $I_{AB}^{(a)}$ of the latter is approximately 1.16 times larger than the former across atom pairs. We do observe that Slater determinant wave functions (DFT and HF, upper row) exaggerate differences between single-bond atom pairs with respect to DMRG. The differentiation is between atom pairs involving an H atom (left cluster) and those that do not (right cluster). So, as can be expected, the high-quality DMRG wavefunction consistently identifies a higher correlation, derived from MI, between atoms, however, the self-consistent field approaches catch the same trends. This observation opens interesting perspectives from a practitioner's point of view when correlations, derived from MI, in peptide systems need to be characterized

at a high quantitative level.

A further point of scrutiny is reserved for the effectiveness of the pragmatic summation $I_{AB} \approx I_{AB}^{(a)}$ in eq. (2.8). For this purpose, we compared $I_{AB}^{(a)}$ with exact block-wise MI measures ($I_{AB}^{(b)}$) between full blocks of AOs associated to certain atoms in the context of DMRG. These blocks represent larger subsystems (e.g., atoms) of the molecule, allowing for MI between constituent atoms to be captured. As dipeptide systems are out of computational reach for this study, we focused on the HCN cyanide molecule, which serves as a small yet informative model for testing $I_{AB}^{(a)}$ against $I_{AB}^{(b)}$.

DMRG calculations were performed with a bond dimension of 4096 and a modified 6-31G basis set. In a first run, we ordered the orbitals such that the entanglement entropy could be computed between carbon's orbitals (left block) and the remaining NH orbitals (right block). In the subsequent DMRG run, the orbitals were reordered to place nitrogen in the left block to measure MI between nitrogen and the remaining system (CH block). After several reordering, $I_{AB}^{(b)}$ was computed for all atomic connections. Here, we remark that the same information could technically be extracted from a single DMRG run by forming proper traces of subsystems using the obtained MPS wave function, however this becomes computational more demanding for large DMRG bond dimension values (see Ref. [51]).

Figure 2.11 presents a detailed comparison between $I_{AB}^{(b)}$ (using DMRG) and $I_{AB}^{(a)}$ (using DMRG, DFT, and RHF). The figure illustrates the evolution of MI as the hydrogen atom dissociates from the cyanide molecule. By comparing the highly accurate block-wise MI with the more computationally efficient summation method, it is possible to evaluate the effectiveness of summation-based approach of 2.8 during hydrogen dissociation.

The results demonstrate that the summation method captures key trends in MI behavior with a reasonable degree of accuracy. This shows that $I_{AB}^{(a)}$ derived from the summation approach reliably approximates the qualitative MI between atoms for

such weakly correlated systems, making it a practical solution for larger molecular systems where full DMRG calculations are computationally prohibitive. For a more rigorous and computationally demanding theory on multi-orbital correlations based on concepts of k-partitionability and k-produciblity, we guide interested readers to Ref. [51].

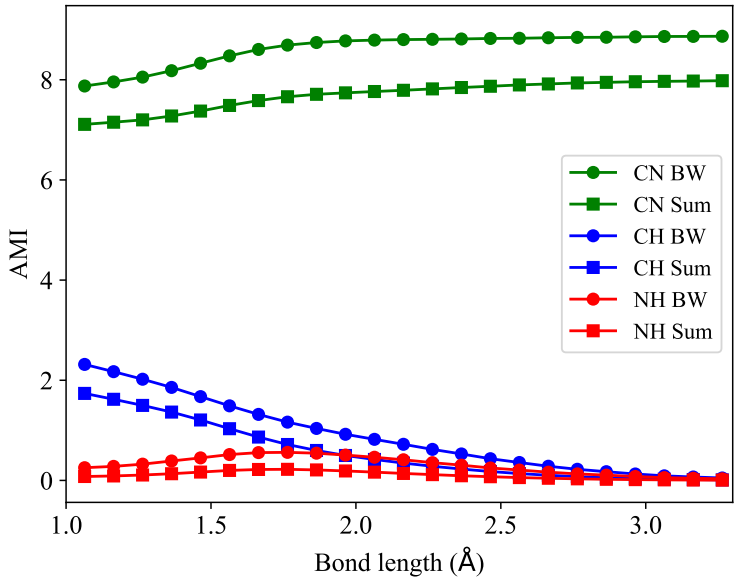


Figure 2.11: Evolution of block-wise (BW, squares) $I^{(b)}_{AB}$ and summation-wise (Sum, dots) $I^{(a)}_{AB}$ AMI for all atom pairwise interactions C-H (blue), C-N (green), and N-H (red) during the hydrogen dissociation in a cyanide system at the DMRG level.

We extended the analysis to compare $I^{(b)}_{AB}$ and $I^{(a)}_{AB}$ in the formaldehyde system. Figure 2.12 displays SPAWN plots for both cyanide and formaldehyde systems, illustrating the MI between atoms, calculated at the DMRG level of theory. Figure 2.12(a) and (c) represent block-wise MI $I^{(b)}_{AB}$, whereas (b) and (d) highlight the difference between the block-wise $I^{(b)}_{AB}$ and summation based $I^{(a)}_{AB}$. The relative small errors in (b) and (d) underline the effectiveness of the summation approach to capture the key atomic interactions and quantum correlations, derived from MI, at a strongly reduced computational cost.

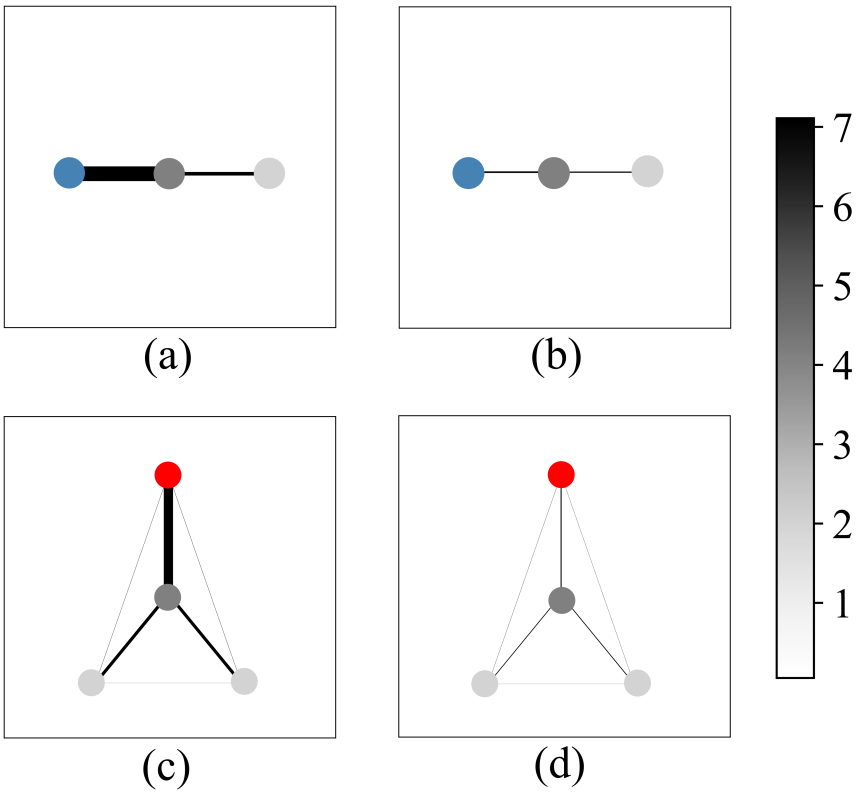
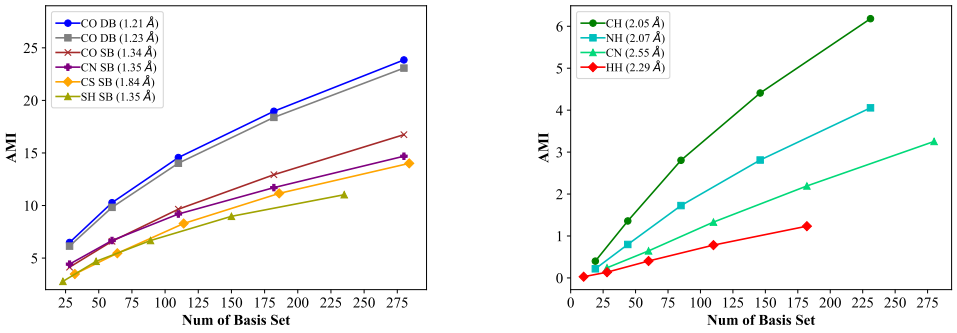


Figure 2.12: SPAWN plots for the cyanide (a) and formaldehyde (c) systems, based on block-wise $I^{(b)}_{AB}$ from DMRG calculations. Panels (b) and (d) display the difference between MI calculated using the block-wise and summation-based approach. The line thickness indicates the strength of the MI between atoms; Carbon, Nitrogen, Oxygen and Hydrogen atoms are represented by black, blue, red and gray dots respectively.

A final note concerns the basis dependence and absolute scale of the presented AMI. As it is challenging to associate a direct experimental observable to AMI, it is difficult to assign meaning to their absolute values. Nevertheless, it can be intuitively understood that entanglement and MI measures will be proportional to the size of the Hilbert space. For this reason, we investigate the impact of basis set size on $I^{(a)}_{AB}$, by analyzing the CysGly dipeptide system using the correlation-consistent polarized valence X-Zeta (cc-pVXZ) basis set series. The cc-pVXZ basis sets were chosen for their systematic and consistent expansion of basis functions, which allowed us to observe how $I^{(a)}_{AB}$ scales with basis set size for different types of interactions, including double bonds, single bonds, and non-bonded interactions. As the number of basis functions

increases, the electronic structure is described with higher spatial resolution, leading to a proportional increase in information content and therefore MI values. The size of the basis set for pairwise interatomic $I_{AB}^{(a)}$ has been chosen as $L_A + L_B$, with L_A and L_B the respective basis set sizes both atom involved in the AMI. As shown in Figure 2.13, $I_{AB}^{(a)}$ values consistently grow linearly with basis set size, pointing out that relative $I_{AB}^{(a)}$ measures of correlation are robust against differing sizes of basis sets.



(a) (b)

Figure 2.13: Effect of increasing basis set size on $I_{AB}^{(a)}$ for the CysGly dipeptide system, using the cc-pVXZ basis set series. The $I_{AB}^{(a)}$ values are plotted for selected (a) covalently bonded and (b) non-bonded atom pairs. While $I_{AB}^{(a)}$ initially increases with basis set size, the growth progressively slows. This nonlinear behavior suggests that AMI captures genuine physical correlation and does not scale indefinitely with basis set size and addresses concerns about potential artifacts due to basis set expansion.

2.4 Conclusion

We investigated the locality of electron correlation, derived from MI, in organic systems through MI indices between AOs and full atomic systems. While many well-established quantum chemical methods exist for large molecules, our approach leverages AMI data as a means to quantify and analyze electronic correlations at the atomic level, providing a complementary perspective that does not rely on predefined bonding models and can be systematically scaled to larger polypeptide systems. More in particular, it can provide a quantitative means for an unbiased partitioning of ex-

tensive systems into smaller fragments in embedding approaches [43, 90, 107]. On the one hand, we confirmed and quantified the expected notion of locality in dipeptide systems through the observation that quantum information correlations decay exponentially with the interatomic distance between atom pairs in dipeptides. On the other hand, we also observed that $I_{AB}^{(a)}$ indices respect chemical notions of bonding, with a hierarchy in $I_{AB}^{(a)}$, highest in double bonds, followed by single bonds, and lowest in non-bonded interactions, highlighting the varying strengths of these chemical interactions. We also demonstrated that self-consistent DFT and HF methods effectively capture electronic correlation trends, in an information theoretic perspective, compared to high-quality and computationally expensive DMRG results, reconfirming the applicability of these methods in the description of weakly correlated and near-equilibrium geometries like dipeptides. Although DFT and HF methods systematically undervalue DMRG results with a factor of 1.16, the general and relative trends are robustly reproduced by these computationally inexpensive methods.

For future work, we envision applying our approach to investigate spatial information theory-based correlations in extensive organic systems, such as proteins, nucleic acids, and other biomolecules. By integrating AMI with conventional chemical analysis methods such as NBO analysis, QTAIM, and EDA, we aim to achieve a more comprehensive understanding of electronic structure and bonding.

The ability to use coarse-grained MI to understand the structure of biomolecules underscores its utility in revealing how chemical groups interact and the strength of these interactions. As research in this area progresses, further refinements and applications of MI are expected to advance our understanding of molecular interactions and facilitate breakthroughs in various scientific fields.

2.5 Acknowledgements

S.D.B acknowledges the NSERC Canada Research Chair and Discovery Grant program for financial support. M.J.M gratefully acknowledges a MITACS Globalink grant to allow for an extensive research stay at the Nicolaus Copernicus University in Toruń. Funded/Co-funded by the European Union (ERC, DRESSED-pCCD, 101077420). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. P.T. is grateful for an OPUS 17 research grant of the National Science Centre, Poland (no. 2019/33/B/ST4/02114). Ö.L. acknowledges support from the Hungarian National Research, Development and Innovation Office (NKFIH) through Grant Nos. K134983 and TKP2021-NVA-04 by the Quantum Information National Laboratory of Hungary, from the Hans Fischer Senior Fellowship programme funded by the Technical University of Munich – Institute for Advanced Study, and from the Center for Scalable and Predictive methods for Excitation and Correlated phenomena, funded as part of the Computational Chemical Sciences Program FWP 70942, by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences at Pacific Northwest National Laboratory.

2.6 Appendix

2.6.1 Mutual Information from Slater determinant wavefunctions

The major benefit of employing Slater determinant (SD) wavefunctions when computing entanglement and correlation measures in extensive systems is that all n-RDMs

reduce to combinations of 1-RDMs. Although it is straightforward to derive this reduction for each RDM involved in the computation of the MI indices [59], it is desirable to have a systematic approach. This approach is provided by Wick's theorem, which allows for the rewriting of each operator into a normal ordering of the involved single-particle operators with respect to the defined vacuum state, corrected by all possible operator contractions, which are equivalent to the 1-RDMs [85, 108].

Wick's theorem reduces the complexity of multi-operator products by breaking down a complex product into a sum over pairwise contractions. In the context of mean field theory, certain contractions vanish because of the properties of the underlying state. Thus, when expanding the four-operator product using Wick's theorem, only the non-vanishing contractions contribute to the final result, which simplifies the calculation considerably. We refer to refs [85, 109] for an in-depth discussion.

To illustrate the procedure, we present the computation of the following matrix element using proper fermionic tensor algebra [110, 59]

$$\langle n_{i\uparrow} c_{i\downarrow} (1 - n_{j\uparrow}) c^\dagger_{j\downarrow} \rangle = \langle c^\dagger_{i\uparrow} c_{i\uparrow} c_{i\downarrow} c^\dagger_{j\downarrow} \rangle - \langle c^\dagger_{i\uparrow} c_{i\uparrow} c_{i\downarrow} c^\dagger_{j\uparrow} c_{j\uparrow} c^\dagger_{j\downarrow} \rangle, \tag{2.9}$$

$$= -\langle c^\dagger_{i\uparrow} c_{i\uparrow} \rangle \langle c^\dagger_{j\downarrow} c_{i\downarrow} \rangle + \langle c^\dagger_{i\uparrow} c_{i\uparrow} \rangle \langle c^\dagger_{j\uparrow} c_{j\uparrow} \rangle \langle c^\dagger_{j\downarrow} c_{i\downarrow} \rangle \; - \langle c^\dagger_{i\uparrow} c_{j\uparrow} \rangle \langle c^\dagger_{j\uparrow} c_{i\uparrow} \rangle \langle c^\dagger_{j\downarrow} c_{i\downarrow} \rangle, \tag{2.10}$$

in which only non-vanishing 1-RDMs due to particle-number and spin-symmetry conservation are retained in the result. The 1-RDMS $\langle c^\dagger_{i\sigma} c_{j\sigma} \rangle$ depend on the choice of orthonormal orbitals. Here, we chose to work with symmetric orthogonalized orbitals (2.7), leading to the convenient expressions

$$\langle c^\dagger_{i\sigma} c_{j\sigma} \rangle = \sum_{\alpha \in \text{occ}} (S^{\frac{1}{2}} A)_{i\alpha} (A^T S^{\frac{1}{2}})_{\alpha j}, \tag{2.11}$$

in which A is the matrix expressing the canonical orbitals of the SD in the basis of the AOs, and the summation over α runs only over the occupied (occ) orbitals in the

SD.

2.6.2 DMRG Calculations

To further validate the use of single-slater determinant methods (DFT and HF) for capturing MI, we performed a set of DMRG calculations for the second dipeptide, AlaGly. Figure 2.14 shows the SPAWN plot for AlaGly derived from DMRG calculations with bond dimension $D = 2560$.

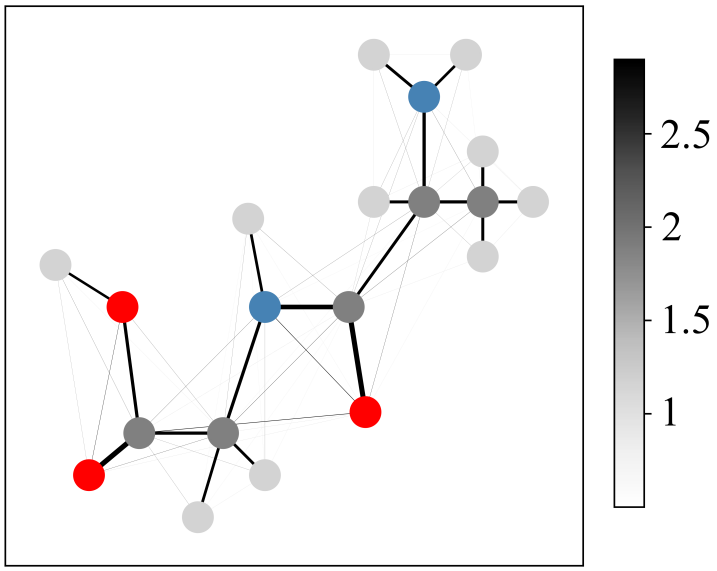
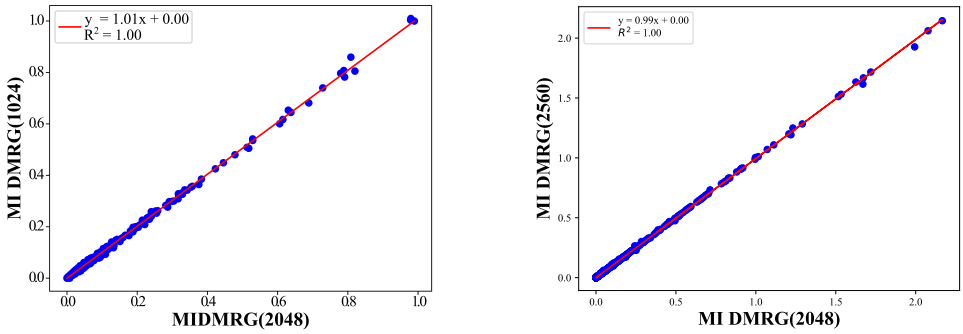


Figure 2.14: The $I_{AB}^{(a)}$ plot for Ala-Gly dipeptide obtained from DMRG ($D = 2560$)

It is important to investigate the convergence of the DMRG computations. For this, computations for different bond dimensions D were performed for GlyGly ($D = 1024, 2048$) and AlaGly ($D = 1024, 2048, 2560$). The near-perfect linear fit in Figure 2.15 suggests that increasing the accuracy of calculations by raising the value of D does not lead to significant changes in the MI, implying that the calculations of the MI distribution are converged.



(a) (b)

Figure 2.15: MI derived from DMRG calculations with different levels of accuracy for (a) GlyGly ($D = 1024, 2048$) and (b) AlaGly ($D = 1024, 2048, 2560$) align with each other.

The DMRG computation for AlaGly ($D = 2560$) required 9078 CPU hours and 400 GB memory, whereas GlyGly ($D = 2048$) needed 7977 CPU hours with 150 GB memory. The DFT and HF computations are generally completed within minutes on a 10-core node.

To further examine the basis dependence of AMI, we constructed a small one-dimensional model system: two electrons confined in a double-well potential. By systematically increasing the number of grid points from 2 to 90, we effectively refined the resolution of the Hilbert space. As shown in Figure 2.16, MI increases initially with grid resolution, reflecting the growing capacity of the basis to describe electronic correlation. However, beyond approximately 70 grid points, the MI curve begins to plateau, indicating convergence to a finite value as the basis becomes sufficiently complete. This saturation behavior supports the physical interpretability of MI when a complete basis is used. These findings are consistent with the framework proposed by Aliverti-Piuri et al. [111], who distinguish between two types of correlation, particle correlation (basis-invariant) and orbital correlation (basis-dependent).

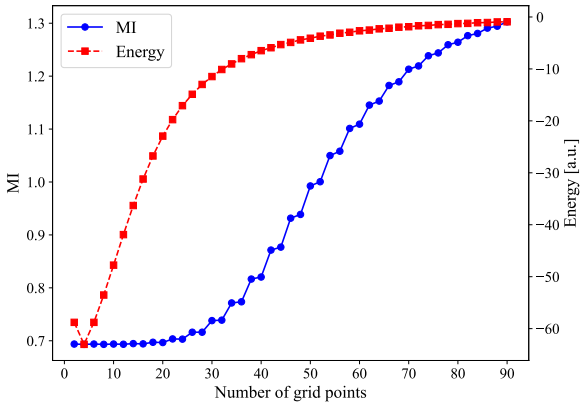


Figure 2.16: Evolution of MI (blue circles, left axis) and total electronic energy (red squares, right axis) as a function of grid resolution in a one-dimensional system with two potential wells and two electrons. As grid resolution increases, the total energy rapidly converges, while MI initially increases before plateauing beyond 70 grid points.

To assess the impact of orbital representation on AMI, we computed AMI using two types of orbital bases: Pipek–Mezey LMOs and orbitals obtained from orbital-optimized [112, 113] pair coupled cluster doubles (OO-pCCD) for the GlyGly dipeptide.

To enable a consistent atom-wise mapping across different orbital sets, we employed a projection-based orbital assignment scheme. Each MO ψ_j was assigned to the atom whose AO basis functions contributed the largest portion of its norm. Specifically, the weight $W_A^{(j)}$ of atom A in orbital ψ_j is defined as:

$$W_A^{(j)} = \sum_{\mu \in A} |C_{\mu j}|^2, \tag{2.12}$$

where $C_{\mu j}$ denotes the coefficient of AO μ in the expansion of MO ψ_j, and the sum is taken over all AOs centered on atom A. Each MO was then assigned to the atom with the highest weight:

$$A^* = \arg\max_A \left(W_A^{(j)} \right). \tag{2.13}$$

The comparison between AO, Pipek–Mezey, and OO-pCCD-based AMI is presented in Figure 2.17, where we show both circular (top row) and SPAWN plot (bottom row) visualizations. The MI values in the MO-based frameworks (Pipek–Mezey and OO-pCCD) are computed by the PyBEST program. [99, 100] This highlights a key difference: AO-based MI values are similar to those obtained from MO bases. All three representations consistently identify strong correlations within covalent bonds with a notion of double and single bonds.

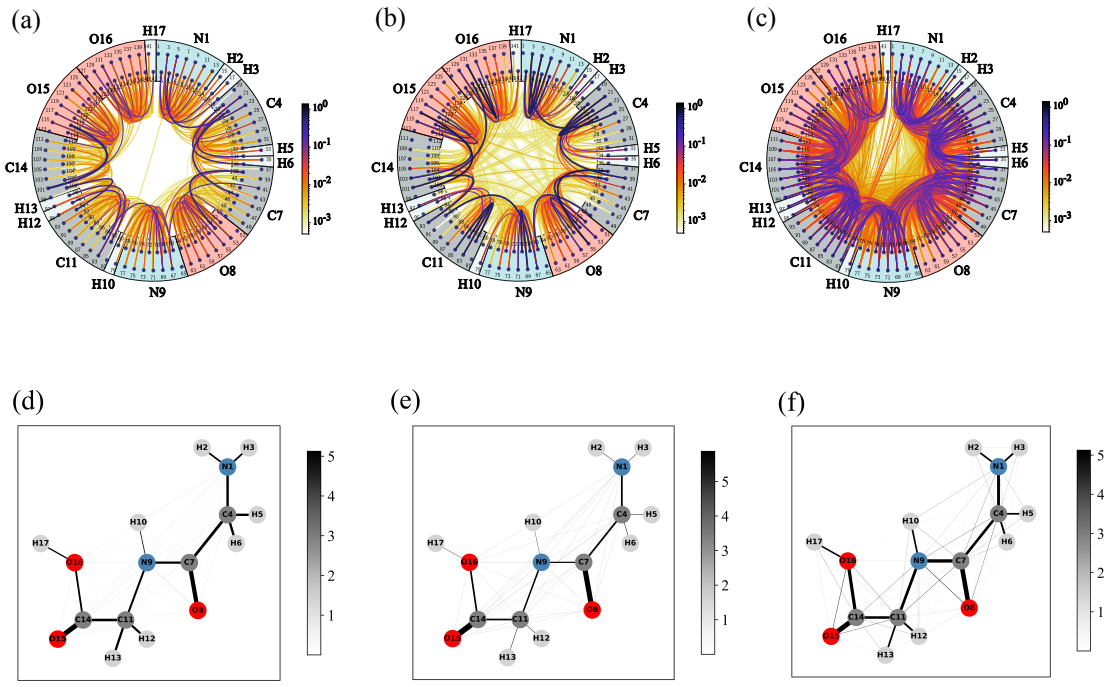


Figure 2.17: (a) Pipek–Mezey localized orbitals circular MI plot, (b) OO-pCCD orbitals circular MI plot, (c) AO circular MI plot for Glycine-Glycine dipeptide. (d–f) Corresponding SPAWN plots for each orbital basis. The color bar in circular plots shows the magnitude of MI on a log scale; the SPAWN plots use edge thickness and grayscale to indicate MI magnitude.

Bibliography

[1] B. Alberts, A. Johnson, J. Lewis, P. Walter, M. Raff, K. Roberts, Molecular Biology of the Cell 4th Edition: International Student Edition, Routledge, 2002.

[2] C. B. Anfinsen, Principles That Govern the Folding of Protein Chains, Sci. 181 (4096) (1973) 223–230.

[3] B.-G. Ma, J.-X. Guo, H.-Y. Zhang, Direct Correlation Between Proteins' Folding Rates and Their Amino Acid Compositions: An Ab Initio Folding Rate Prediction, Proteins: Struct. Funct. Bioinf. 65 (2) (2006) 362–372.

[4] M. Born, J. R. Oppenheimer, Zur Quantentheorie Der Molekeln, Ann. Phys. 84 (1927) 457–484.

[5] T. Helgaker, P. Jorgensen, J. Olsen, Molecular Electronic-Structure Theory, John Wiley & Sons, 2013.

[6] S. Bagheri, H. Behnejad, R. Firouzi, M. H. Karimi-Jafari, Using the Semiempirical Quantum Mechanics in Improving the Molecular Docking: A Case Study With CDK2, Mol. Inf. 39 (9) (2020) 2000036.

[7] M. Lepšík, J. Řezáč, M. Kolář, A. Pecina, P. Hobza, J. Fanfrlík, The Semiempirical Quantum Mechanical Scoring Function for in Silico Drug Design, Comput. Phys. Commun. 78 (9) (2013) 921–931.

[8] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, J. Am. Chem. Soc. 118 (45) (1996) 11225–11236.

[9] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, J. Comput. Chem. 4 (2) (1983) 187–217.

[10] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, et al., Highly Accurate Protein Structure Prediction With AlphaFold, Nature 596 (7873) (2021) 583–589.

[11] S.-L. J. Lahey, C. N. Rowley, Simulating Protein–Ligand Binding With Neural Network Potentials, Chem. Sci. 11 (9) (2020) 2362–2368.

[12] X. Periole, S.-J. Marrink, The Martini Coarse-Grained Force Field, Biomol. Simul.: Methods Protocols (2013) 533–565.

[13] M. Baaden, S. J. Marrink, Coarse-Grain Modelling of Protein–protein Interactions, Curr. Opin. Struct. Biol. 23 (6) (2013) 878–886.

[14] N. Singh, W. Li, Recent Advances in Coarse-Grained Models for Biomolecules and Their Applications, Int. J. Mol. Sci. 20 (15) (2019) 3774.

[15] K. Berka, R. Laskowski, K. E. Riley, P. Hobza, J. Vondrasek, Representative Amino Acid Side Chain Interactions in Proteins. A Comparison of Highly Accurate Correlated Ab Initio Quantum Chemical and Empirical Potential Procedures, J. Chem. Theory Comput. 5 (4) (2009) 982–992.

[16] A. V. Morozov, K. M. Misura, K. Tsemekhman, D. Baker, Comparison of Quantum Mechanics and Molecular Mechanics Dimerization Energy Landscapes for Pairs of Ring-Containing Amino Acids in Proteins, J. Phys. Chem. B 108 (24) (2004) 8489–8496.

[17] F. M. Bickelhaupt, E. J. Baerends, Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry, Rev. Comput. Chem. (2000) 1–86.

[18] M. v. Hopffgarten, G. Frenking, Energy Decomposition Analysis, WIREs Comput. Mol. Sci. 2 (1) (2012) 43–62.

[19] R. F. W. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, 1990.

[20] E. D. Glendening, C. R. Landis, F. Weinhold, Natural Bond Orbital Methods, WIREs Comput. Mol. Sci. 2 (1) (2012) 1–42.

[21] M. J. Phipps, T. Fox, C. S. Tautermann, C.-K. Skylaris, Energy Decomposition Analysis Approaches and Their Evaluation on Prototypical Protein–Drug Interaction Patterns, Chem. Soc. Rev. 44 (10) (2015) 3177–3211.

[22] D. S. Levine, M. Head-Gordon, Energy Decomposition Analysis of Single Bonds Within Kohn–Sham Density Functional Theory, Proc. Natl. Acad. Sci. 114 (48) (2017) 12649–12656.

[23] P. Popelier, R. Bader, Effect of Twisting a Polypeptide on Its Geometry and Electron Distribution, J. Phys. Chem. 98 (16) (1994) 4473–4481.

[24] C. F. Matta, R. F. Bader, Atoms-in-Molecules Study of the Genetically Encoded Amino Acids. II. Computational Study of Molecular Geometries, Proteins: Struct. Funct. Bioinf. 48 (3) (2002) 519–538.

[25] C. F. Matta, R. F. Bader, An Atoms-In-Molecules Study Of The Genetically-Encoded Amino Acids: I. Effects Of Conformation And Of Tautomerization On Geometric, Atomic, And Bond Properties, Proteins: Struct. Funct. Bioinf. 40 (2) (2000) 310–329.

[26] C. F. Matta, R. F. Bader, Atoms-In-Molecules Study Of The Genetically Encoded Amino Acids. Iii. Bond And Atomic Properties And Their Correlations With Experiment Including Mutation-Induced Changes In Protein Stability And Genetic Coding, Proteins: Struct. Funct. Bioinf. 52 (3) (2003) 360–399.

[27] F. Cortés-Guzmán, R. Bader, Transferability Of Group Energies And Satisfaction Of The Virial Theorem, Chem. Phys. Lett. 379 (1-2) (2003) 183–192.

[28] J. Hernández-Trujillo, R. Bader, Properties Of Atoms In Molecules: Construction Of One-Density Matrix From Functional Group Densities, J. Chem. Phys. 115 (23) (2001) 10595–10607.

[29] M. A. Spackman, The Quantum Theory of Atoms in Molecules. From Solid State to DNA and Drug Design. Edited by Chérif F. Matta and Russell J. Boyd., Angew. Chem. Int. Ed. 46 (36) (2007) 6766–6767.

[30] F. Weinhold, Natural Bond Orbital Methods, WIREs Comput. Mol. Sci. 3 (2002).

[31] M. K. Kesharwani, A. Karton, J. M. Martin, Benchmark Ab Initio Conformational Energies for the Proteinogenic Amino Acids Through Explicitly Correlated Methods. Assessment of Density Functional Methods, J. Chem. Theory Comput. 12 (1) (2016) 444–454.

[32] K. Carter-Fenk, K. U. Lao, J. M. Herbert, Predicting and Understanding Non-Covalent Interactions Using Novel Forms of Symmetry-Adapted Perturbation Theory, Acc. Chem. Res. 54 (19) (2021) 3679–3690.

[33] O. Sinanoğlu, D. F. Tuan, Any-Electron Theory of Atoms and Molecules. III. Effect of Correlation on Orbitals, J. Chem. Phys. 38 (1963) 1740–1748.

[34] T. J. Lee, P. R. Taylor, A Diagnostic for Determining the Quality of Single-Reference Electron Correlation Methods, Int. J. Quantum Chem. 23 (1989) 199–207.

[35] R. J. Bartlett, J. F. Stanton, Applications of Post-Hartree-Fock Methods: A Tutorial, Rev. Comput. Chem. 5 (1994) 165–169.

[36] K. Boguslawski, P. Tecmer, Ö. Legeza, M. Reiher, Entanglement Measures for Single-and Multireference Correlation Effects, J. Phys. Chem. Lett. 3 (21) (2012) 3129–3135.

[37] Y. Mao, M. Loipersberger, P. R. Horn, A. Das, O. Demerdash, D. S. Levine, S. Prasad Veccham, T. Head-Gordon, M. Head-Gordon, From Intermolecular Interaction Energies and Observable Shifts to Component Contributions and Back Again: A Tale of Variational Energy Decomposition Analysis, Annu. Rev. Phys. Chem. 72 (2021) 641–666.

[38] J. Rissler, R. M. Noack, S. R. White, Measuring Orbital Interaction Using Quantum Information Theory, Chem. Phys. 323 (2006) 519–531.

[39] G. Barcza, Ö. Legeza, K. H. Marti, M. Reiher, Quantum-Information Analysis of Electronic States of Different Molecular Structures, Phys. Rev. A 83 (1) (2011) 012508.

[40] E. Fertitta, B. Paulus, G. Barcza, Ö. Legeza, Investigation of Metal–Insulator-Like Transition Through the Ab Initio Density Matrix Renormalization Group Approach, Phys. Rev. B 90 (2014) 245129.

[41] L. Freitag, S. Knecht, S. F. Keller, M. G. Delcey, F. Aquilante, T. B. Pedersen, R. Lindh, M. Reiher, L. Gonzalez, Orbital Entanglement and CASSCF Analysis of the Ru-No Bond in a Ruthenium Nitrosyl Complex, Phys. Chem. Chem. Phys. 17 (2015) 13769–13769.

[42] Y. Kurashige, G. K.-L. Chan, T. Yanai, Entangled Quantum Electronic Wavefunctions of the Mn4CaO5 Cluster in Photosystem II, Nat. Commun. 5 (8) (2013) 660–666.

[43] P. Tecmer, K. Boguslawski, Ö. Legeza, M. Reiher, Unravelling the Quantum-Entanglement Effect of Noble Gas Coordination on the Spin Ground State of CUO, Phys. Chem. Chem. Phys. 16 (2014) 719–727.

[44] K. Boguslawski, P. Tecmer, Orbital Entanglement in Quantum Chemistry, Int. J. Quantum Chem. 115 (19) (2015) 1289–1295.

[45] C. Duperrouzel, P. Tecmer, K. Boguslawski, G. Barcza, Ö. Legeza, P. W. Ayers, A Quantum Informational Approach for Dissecting Chemical Reactions, Chem. Phys. Lett. 621 (2015) 160–164.

[46] P. T. K. Boguslawski, Erratum: Orbital Entanglement in Quantum Chemistry, Int. J. Quantum Chem. 117 (24) (2017) e25455.

[47] P. Tecmer, K. Boguslawski, P. W. Ayers, Singlet Ground State Actinide Chemistry With Geminals, Phys. Chem. Chem. Phys. 17 (2015) 14427–14436.

[48] G. Barcza, R. M. Noack, J. Sólyom, Ö. Legeza, Entanglement Patterns and Generalized Correlation Functions in Quantum Many-Body Systems, Phys. Rev. B 92 (12) (2015) 125140.

[49] Y. Zhao, K. Boguslawski, P. Tecmer, C. Duperrouzel, G. Barcza, Ö. Legeza, P. W. Ayers, Dissecting the Bond-Formation Process of D 10-Metal–ethene Complexes With Multireference Approaches, Theor. Chem. Acc. 134 (10) (2015) 120.

[50] S. Szalay, M. Pfeffer, V. Murg, G. Barcza, F. Verstraete, R. Schneider, Ö. Legeza, Tensor Product Methods and Entanglement Optimization for Ab Initio Quantum Chemistry, Int. J. Quantum Chem. 115 (2015) 1342–1391.

[51] S. Szalay, G. Barcza, T. Szilvási, L. Veis, Ö. Legeza, The Correlation Theory of the Chemical Bond, Sci. Rep. 7 (1) (2017) 2237.

[52] Z. Li, S. Guo, Q. Sun, G. K.-L. Chan, Electronic Landscape of the P-Cluster of Nitrogenase as Revealed Through Many-Electron Quantum Wavefunction Simulations, Nat. Commun. 11 (11) (2019) 1026–1033.

[53] C. Stemmle, B. Paulus, Ö. Legeza, Analysis of Electron-Correlation Effects in Strongly Correlated Systems (N_2 and N_2^+) by Applying the Density-Matrix Renormalization-Group Method and Quantum Information Theory, Phys. Rev. A 97 (2) (2018) 022505.

[54] L. Ding, C. Schilling, Correlation Paradox of the Dissociation Limit: A Quantum Information Perspective, J. Chem. Theory Comput. 16 (7) (2020) 4159–4175.

[55] L. Ding, S. Mardazad, S. Das, S. Szalay, U. Schollwoöck, Z. Zimborás, C. Schilling, Concept of Orbital Entanglement and Correlation in Quantum Chemistry, J. Chem. Theory Comput. 17 (1) (2020) 79–95.

[56] A. Leszczyk, T. Dome, P. Tecmer, D. Kędziera, K. Boguslawski, Resolving the Π-Assisted U–N Σ F-Bond Formation Using Quantum Information Theory, Phys. Chem. Chem. Phys. 24 (35) (2022) 21296–21307.

[57] L. Ding, G. Dünnweber, C. Schilling, Physical Entanglement Between Localized Orbitals, Quantum Sci. Technol. 9 (1) (2023) 015005.

[58] M. Gałyńska, M. M. F. de Moraes, P. Tecmer, K. Boguslawski, Delving Into the Catalytic Mechanism of Molybdenum Cofactors: A Novel Coupled Cluster Study, Phys. Chem. Chem. Phys. 26 (27) (2024) 18918–18929.

[59] K. Boguslawski, P. Tecmer, G. Barcza, Ö. Legeza, M. Reiher, Orbital Entanglement in Bond-Formation Processes, J. Chem. Theory Comput. 9 (7) (2013) 2959–2973.

[60] M. Mottet, P. Tecmer, K. Boguslawski, Ö. Legeza, M. Reiher, Quantum Entanglement in Carbon-Carbon, Carbon-Phosphorus and Silicon-Silicon Bonds, Phys. Chem. Chem. Phys. 16 (2014) 8872–8880.

[61] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, Phys. Rev. Lett. 69 (19) (1992) 2863.

[62] S. R. White, R. L. Martin, Ab Initio Quantum Chemistry Using the Density Matrix Renormalization Group, J. Chem. Phys. 110 (9) (1999) 4127–4130.

[63] S. R. White, Density-Matrix Algorithms for Quantum Renormalization Groups, Phys. Rev. B 48 (1993) 10345–10356.

[64] U. Schollwöck, The Density-Matrix Renormalization Group in the Age of Matrix Product States, Ann. Phys. 326 (1) (2011) 96–192.

[65] U. Schollwöck, The Density-Matrix Renormalization Group, Rev. Mod. Phys. 77 (2005) 259–315.

[66] Ö. Legeza, J. Sólyom, Optimizing the Density-Matrix Renormalization Group Method Using Quantum Information Entropy, Phys. Rev. B 68 (2003) 195116.

[67] T. Yanai, Y. Kurashige, E. Neuscamman, G. K.-L. Chan, Multireference Quantum Chemistry Through a Joint Density Matrix Renormalization Group and Canonical Transformation Theory, J. Chem. Phys. 132 (2) (2010) 024105.

[68] S. Wouters, D. Van Neck, The Density Matrix Renormalization Group for Ab Initio Quantum Chemistry, Eur. Phys. J. D 68 (2014) 1–20.

[69] K. H. Marti, I. M. Ondík, G. Moritz, M. Reiher, Density Matrix Renormalization Group Calculations on Relative Energies of Transition Metal Complexes and Clusters, J. Chem. Phys. 128 (1) (2008) 014104.

[70] A. Menczer, M. van Damme, A. Rask, L. Huntington, J. Hammond, S. S. Xantheas, M. Ganahl, Ö. Legeza, Parallel Implementation of the Density Matrix Renormalization Group Method Achieving a Quarter petaFLOPS Performance on a Single DGX-H100 GPU Node, J. Chem. Theory Comput. 20 (19) (2024) 8397–8404.

[71] Ö. Legeza, R. M. Noack, J. Sólyom, L. Tincani, Applications of Quantum Information in the Density-Matrix Renormalization Group, in: Computational Many-Particle Physics, Vol. 739 of Lect. Notes Phys., Springer, Berlin/Heidelerg, 2008, pp. 653–664.

[72] S. Wouters, W. Poelmans, P. W. Ayers, D. Van Neck, CheMPS2: A Free Open-Source Spin-Adapted Implementation of the Density Matrix Renormalization Group for Ab Initio Quantum Chemistry, Comput. Phys. Commun. 185 (6) (2014) 1501–1514.

[73] C. Krumnow, L. Veis, Ö. Legeza, J. Eisert, Fermionic Orbital Optimization in Tensor Network States, Phys. Rev. Lett. 117 (21) (2016) 210402.

[74] C. Krumnow, L. Veis, J. Eisert, Ö. Legeza, Effective Dimension Reduction With Mode Transformations: Simulating Two-Dimensional Fermionic Condensed Matter Systems With Matrix-Product States, Phys. Rev. B 104 (7) (2021) 075137.

[75] M. Máté, K. Petrov, S. Szalay, Ö. Legeza, Compressing Multireference Character of Wave Functions via Fermionic Mode Optimization, J. Math. Chem. 61 (2) (2023) 362–375.

[76] W. Mizukami, Y. Kurashige, T. Yanai, More Π Electrons Make a Difference: Emergence of Many Radicals on Graphene Nanoribbons Studied by Ab Initio DMRG Theory, J. Chem. Theory Comput. 9 (1) (2013) 401–407.

[77] K. Boguslawski, P. Tecmer, Ö. Legeza, Analysis of Two-Orbital Correlations in Wave Functions Restricted to Electron-Pair States, Phys. Rev. B 94 (15) (2016) 155126.

[78] J. Foster, S. Boys, Canonical Configurational Interaction Procedure, Rev. Mod. Phys. 32 (1960) 300.

[79] J. Pipek, P. G. Mezey, A Fast Intrinsic Localization Procedure Applicable for Abinitio and Semiempirical Linear Combination of Atomic Orbital Wave Functions, J. Chem. Phys. 90 (1989) 4916–4926.

[80] D. Pekker, C. Liang, S. Pattanayak, S. Mukhopadhyay, Machine Learning 1- And 2-Electron Reduced Density Matrices of Polymeric Molecules (2022). arXiv:2208.04976.

[81] J. J. P. Stewart, An Examination of the Nature of Localized Molecular Orbitals and Their Value in Understanding Various Phenomena That Occur in Organic Chemistry, J. Mol. Model. 25 (2019) 7.

[82] S. Jahani, K. Boguslawski, P. Tecmer, The Relationship Between Structure and Excited-State Properties in Polyanilines From Geminal-Based Methods, RSC Adv. 13 (40) (2023) 27898–27911.

[83] A. Nowak, Ö. Legeza, K. Boguslawski, Orbital Entanglement and Correlation From pCCD-tailored Coupled Cluster Wave Functions, J. Chem. Phys. 154 (8) (2021) 084111.

[84] D. Materia, L. Ratini, C. Angeli, L. Guidoni, Quantum Information Reveals That Orbital-Wise Correlation Is Essentially Classical in Natural Orbitals, J. Chem. Phys. 161 (24) (2024) 244108.

[85] G. C. Wick, The Evaluation of the Collision Matrix, Phys. Rev. 80 (1950) 268–272.

[86] A. Karton, P. R. Spackman, Evaluation of Density Functional Theory for a Large and Diverse Set of Organic and Inorganic Equilibrium Structures, J. Comput. Chem. 42 (22) (2021) 1590–1601.

[87] V. K. Prasad, Z. Pei, S. Edelmann, A. Otero-de-la Roza, G. A. DiLabio, BH9, a New Comprehensive Benchmark Data Set for Barrier Heights and Reaction Energies: Assessment of Density Functional Approximations and Basis Set Incompleteness Potentials, J. Chem. Theory Comput. 18 (1) (2022) 151–166.

[88] P. C. St. John, Y. Guan, Y. Kim, S. Kim, R. S. Paton, Prediction of Organic Homolytic Bond Dissociation Enthalpies at Near Chemical Accuracy With Sub-Second Computational Cost, Nat. Commun. 11 (1) (2020) 2328.

[89] A. Szabo, N. S. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Courier Corporation, 1996.

[90] R. Chakraborty, K. Boguslawski, P. Tecmer, Static Embedding With Pair Coupled Cluster Doubles Based Methods, Phys. Chem. Chem. Phys. 25 (2023) 25377–25388.

[91] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields, J. Phys. Chem. 98 (45) (1994) 11623–11627.

[92] N. Mardirossian, M. Head-Gordon, Mapping the Genome of Meta-Generalized Gradient Approximation Density Functionals: The Search for B97m-V, J. Chem. Phys. 142 (7) (2015) 074111.

[93] N. Mardirossian, M. Head-Gordon, Thirty Years of Density Functional Theory in Computational Chemistry: An Overview and Extensive Assessment of 200 Density Functionals, Mol. Phys. 115 (19) (2017) 2315–2372.

[94] Ö. Legeza, L. Veis, T. Mosoni, QC-DMRG-Budapest, a Program for Quantum Chemical DMRG Calculations (2000).

[95] F. Neese, The ORCA Program System, WIREs Comput. Mol. Sci. 2 (1) (2012) 73–78.

[96] C. J. Stein, M. Reiher, Automated Selection of Active Orbital Spaces, J. Chem. Theory Comput. 12 (4) (2016) 1760–1771.

[97] C. J. Stein, D. A. Pantazis, V. Krewald, Orbital Entanglement Analysis of Exchange-Coupled Systems, J. Phys. Chem. Lett. 10 (21) (2019) 6762–6770.

[98] I. Hagymási, Ö. Legeza, Entanglement, Excitations, and Correlation Effects in Narrow Zigzag Graphene Nanoribbons, Phys. Rev. B 94 (16) (2016) 165147.

[99] K. Boguslawski, A. Leszczyk, A. Nowak, F. Brzęk, P. S. Żuchowski, D. Kędziera, P. Tecmer, Pythonic Black-Box Electronic Structure Tool (PyBEST). An Open-Source Python Platform for Electronic Structure Calculations at the Interface Between Chemistry and Physics, Comput. Phys. Commun. 264 (2021) 107933.

[100] K. Boguslawski, F. Brzęk, R. Chakraborty, K. Cieślak, S. Jahani, A. Leszczyk, A. Nowak, E. Sujkowski, J. Świerczyński, S. Ahmadkhani, D. Kędziera, M. H. Kriebel, P. S. Żuchowski, P. Tecmer, PyBEST: Improved Functionality and Enhanced Performance, Comput. Phys. Commun. 297 (2024) 109049.

[101] D. Koch, M. Pavanello, X. Shao, M. Ihara, P. W. Ayers, C. F. Matta, S. Jenkins, S. Manzhos, The Analysis of Electron Densities: From Basics to Emergent Applications, Chem. Rev. 124 (22) (2024) 12661–12737.

[102] C. F. Matta, P. W. Ayers, R. Cook, Electron Localization-Delocalization Matrices, Springer, 2024.

[103] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese, Sparse Maps—A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory, J. Chem. Phys. 144 (2) (2016) 024109.

[104] P. R. Nagy, M. Kállay, Approaching the Basis Set Limit of CCSD (T) Energies for Large Molecules With Local Natural Orbital Coupled-Cluster Methods, J. Chem. Theory Comput. 15 (10) (2019) 5275–5298.

[105] L. Gyevi-Nagy, M. Kállay, P. R. Nagy, Integral-Direct and Parallel Implementation of the CCSD(T) Method: Algorithmic Developments and Large-Scale Applications, J. Chem. Theory Comput. 16 (1) (2019) 366–384.

[106] P. B. Szabó, J. Csóka, M. Kállay, P. R. Nagy, Linear-Scaling Local Natural Orbital CCSD(T) Approach for Open-Shell Systems: Algorithms, Benchmarks, and Large-Scale Applications, J. Chem. Theory Comput. 19 (22) (2023) 8166–8188.

[107] G. Knizia, G. K.-L. Chan, Density Matrix Embedding: A Simple Alternative to Dynamical Mean-Field Theory, Phys. Rev. Lett. 109 (2012) 186404.

[108] P. Ring, P. Schuck, The Nuclear Many-Body Problem, 3rd Edition, Springer, Berlin, 2004.

[109] A. L. Fetter, J. D. Walecka, Quantum Theory of Many-Particle Systems, Courier Corporation, 2012.

[110] S. Szalay, Z. Zimborás, M. Máté, G. Barcza, C. Schilling, Ö. Legeza, Fermionic Systems for Quantum Information People, J. Phys A-Mat. Theor. 54 (39) (2021) 393001.

[111] D. Aliverti-Piuri, K. Chatterjee, L. Ding, K. Liao, J. Liebert, C. Schilling, What Can Quantum Information Theory Offer to Quantum Chemistry?, Faraday Discuss. 254 (2024) 76–106.

[112] K. Boguslawski, P. Tecmer, P. W. Ayers, P. Bultinck, S. De Baerdemacker, D. Van Neck, Efficient Description of Strongly Correlated Electrons, Phys. Rev. B 89 (2014) 201106(R).

[113] K. Boguslawski, P. Tecmer, P. W. Ayers, P. Bultinck, S. De Baerdemacker, D. Van Neck, Non-Variational Orbital Optimization Techniques for theAP1roG Wave Function, J. Chem. Theory Comput. 10 (2014) 4873–4882.

Chapter 3

From dipeptide systems to polypeptides: evolution of mutual information.

Mostafa Javaheri Moghadam[1], **Katharina Boguslawski**[2], **Paweł Tecmer**[2], **Stijn De Baerdemacker**[1,3]

[1]Department of Chemistry, University of New Brunswick, Fredericton, NB, Canada

[2]Institute of Physics, Faculty of Physics, Astronomy, and Informatics, Nicolaus Copernicus University in Toruń, Grudziądzka 5, 87-100 Toruń, Poland

[3]Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, Canada.

Abstract

Understanding the electronic structure of amino acids is crucial for understanding protein stability, folding mechanisms, and molecular interactions. In this study, we introduce Fragment-wise Mutual Information (FMI) as a quantum information-based tool to quantify inter-fragment correlations in peptides. By extending mutual information (MI) analysis to amino acid fragments, FMI provides a detailed map of electronic interactions beyond classical descriptors such as van der Waals forces. We first validate FMI on 400 dipeptides, demonstrating a correlation with atomization and bonding energies. Expanding this approach to the 10-mer Neh2 peptide, we analyze molecular dynamics (MD) simulations and reveal how electronic correlations evolve during folding. Our results show that FMI distinguishes stabilizing interactions such as salt bridges and variable hydrogen bond strengths, providing deeper insight into peptide stability. These findings suggest that FMI could enhance molecular modeling and force field development by incorporating quantum electronic effects into biomolecular analysis.

3.1 Introduction

Electronic interactions govern the structural stability and reactivity of biomolecules. Accurately capturing these interactions is essential for understanding complex biological processes such as protein folding, ligand binding, and enzymatic activity. However, the complexity of electronic correlations makes it challenging to model these effects accurately.

High-quality quantum mechanical methods such as the density matrix renormalization group (DMRG) algorithm [1, 2, 3, 4, 5, 6], coupled cluster (CC) [7, 8, 9] and density functional theory (DFT) [10, 11] provide accurate descriptions of electronic structures. While DMRG and CC offer systematically improvable treatments of electron

correlation, their high computational cost limits their applicability to large biomolecular systems. While DFT is not systematically improvable like DMRG and CC, it offers more computationally traceable results with reasonable accuracy. However, as biomolecular systems grow in size and complexity, the computational demands of these methods quickly become prohibitive.

To address these limitations, empirical and semi-empirical methods have been developed as computationally efficient alternatives [12]. While these approaches offer significant speed advantages, they often rely on parametrization and classical approximations, which can overlook the intricate electronic correlations governing biomolecular stability and interactions. As a result, there remains a pressing need for approaches that strike a balance between accuracy and computational efficiency, enabling the study of large biomolecular systems while preserving key electronic effects [13, 14].

Several quantum-based frameworks have been introduced to analyze molecular interactions and chemical bonding to bridge this gap. The Quantum Theory of Atoms in Molecules (QTAIM), developed by Bader [15], applies topological analysis of the electron density to reveal bonding patterns and delocalization effects. However, its interpretations may be sensitive to user-defined thresholds. Similarly, Natural Bond Orbital (NBO) [14] analysis offers chemically intuitive representations by localizing molecular wavefunctions into classical bonding motifs but may miss complex delocalization and non-covalent effects. These limitations underscore the ongoing need for complementary approaches that capture electron correlation more generally and quantitatively and highlight the need for complementary tools that provide a more general and quantitative measure of electronic interactions.

MI, originally formulated in information theory, quantifies the degree of correlation between two subsystems. In the context of quantum chemistry, MI has been employed to analyze electronic structures, particularly in molecular orbital theory and DMRG studies. This approach has provided valuable insights into the role of electron

correlation in chemical bonding [16, 17, 18, 19, 20, 21, 22, 23, 24].

We recently introduced atomic mutual information (AMI) [25], a novel framework for examining electronic correlations at the atomic level offering a high-resolution perspective on intra and intermolecular interactions. By quantifying the shared electronic information between atomic fragments, AMI allows for a refined understanding of molecular systems beyond conventional energy-based descriptors.

Unlike traditional pairwise interaction descriptors such as hydrogen bonding, van der Waals interactions, and electrostatic forces, AMI captures higher-order dependencies that influence molecular behavior. Leveraging AMI allows for a more resolute insight into the fundamental electronic factors driving molecular interactions, where classical force fields and empirical functions often fail to account for these effects, resulting in limitations in their accuracy.

To explore how electronic correlations contribute to peptide stability and folding, we applied our methodology first to a dataset of 400 dipeptides. Dipeptides were chosen as model systems due to their small size and computational efficiency, while still preserving essential characteristics of peptide bonding and molecular stability. We demonstrated the utility of AMI to explore various electronic correlation patterns and extract statistical trends in our previous work [25]. In the present study, we extend our analysis using a fragmentwise approach, aggregating atomic level correlations into amino acid fragments to calculate FMI. Here, we investigated the application of FMI in characterizing molecular properties of dipeptides and beyond. To examine the relationship between electronic correlation and molecular stability, FMI values were compared with atomization energies, and bonding energies from Energy Decomposition Analysis (EDA) [26, 27], and non-bonded interactions computed using Lennard-Jones potentials [28] with force field parameters from CHARMM [29] and OPLS-AA [30]. These findings served as a foundation for extending our analysis to larger and more complex biomolecular systems.

Beyond dipeptides, we sought to extend our study to larger peptides. In particular, we explored how electronic correlations contribute to the stability and folding of the 10-mer Neh2 peptide. Neh2 is a well-characterized intrinsically disordered protein (IDP) [31, 32, 33], involved in regulating the Nrf2 transcription factor [34, 35]. Nrf2 is a critical component of cellular responses to oxidative stress and Neh2 is a key domain within the Nrf2 protein. Neh2's main function is to mediate the interaction between Nrf2 and Keap1, a regulatory protein that acts as the primary inhibitor of Nrf2 by targeting it for ubiquitination, a process of modifying a protein to regulate its degradation. Unlike structured proteins, IDPs lack a stable tertiary conformation, leading to dynamic and complex electronic interactions. We investigated electronic correlation dynamics in both the native folded structure, Beta-turn, and unfolded states of 10-mer Neh2 using MD simulations followed by DFT calculations at the ωB97M-V/6-31G(d) [36, 37]level. This research aims to provide a new perspective on how electronic correlations evolve and can be resolved during folding and binding.

3.2 Methodology

Our methodology builds upon our prior work [38, 39, 40, 17, 41, 42, 43, 44, 25] in exploring electronic correlations in dipeptide systems. In the present work, we study MI and its coarse-grained analogs to analyze interactions in peptide systems.

3.2.1 Calculation of Mutual Information

Entanglement in a bi-partite quantum system is characterized by the von Neumann entropy derived from the reduced density matrix (RDM). Given a pure state defined over subsystem quantum states $|i\rangle \otimes |e\rangle$

$$|\Psi\rangle = \sum_{i,e} \Psi_{i,e} |i\rangle \otimes |e\rangle, \tag{3.1}$$

the corresponding RDM is obtained by tracing out the degrees of freedom associated with one of the subsystems:

$$\rho = \mathrm{Tr}_{|e\rangle}|\Psi\rangle\langle\Psi| \tag{3.2}$$

If we restrict $|i\rangle$ to be the one-orbital Fock space of the (spatial) orbital i, the above equation defines the orbital-reduced density matrix (ORDM). From the eigenvalues ω_α of ρ, the single-orbital entropy [38] for orbital i is defined as:

$$s(1)_i = -\sum_{\alpha=1}^{4} w_{\alpha,i} \ln w_{\alpha,i}, \tag{3.3}$$

which quantifies the entanglement of a single orbital with the rest of the system. Since for spatial orbitals, the one-orbital Fock space is spanned by four possible states (unoccupied, singly occupied, doubly occupied), each ORDM ρ has four eigenvalues. The two-orbital entropy extends this to orbital pairs (i, j) [38]:

$$s(2)_{ij} = -\sum_{\alpha=1}^{16} w_{\alpha,ij} \ln w_{\alpha,ij}. \tag{3.4}$$

where we have 16 different occupations in the two-orbital Fock space (see also [38, 41]). MI is then computed as:

$$I_{ij} = s(2)_{ij} - s(1)_i - s(1)_j, \tag{3.5}$$

which quantifies both quantum and classical correlations between orbitals i and j. This measure enables the identification of key interactions within the molecular framework.

As before, we utilized atomic orbitals (AOs) and applied symmetric orthogonalization to account for nonzero overlaps:

$$c_{j\sigma}^{\dagger} = \sum_{i=1}^{L} S_{ij}^{-\frac{1}{2}} a_{i\sigma}^{\dagger} \tag{3.6}$$

where S is the AO overlap matrix and $a^{\dagger}$ is the fermionic creation operator for spin orbital $i\sigma$. This transformation ensures an orthonormal basis while preserving the locality of orbitals.

To derive chemically meaningful descriptors at the atom-pair level, we coarse-grained MI from the orbital level to the atomic level. This is achieved by summing all pairwise MI contributions within a given atom

$$I_{AB} \approx I_{AB}^{(a)} = \sum_{i=1}^{L_A} \sum_{j=1}^{L_B} I_{ij}, \tag{3.7}$$

where A and B denote atoms in the system and L_A and L_B represent the total number of spatial orbitals in atoms A and B, respectively. We refer to the $I_{AB}^{(a)}$ as the (approximate) AMI. To extend this framework further, FMI is introduced by treating amino acid residues as fundamental units:

$$I_{XY} \approx \sum_{A \in X} \sum_{B \in Y} I_{AB}^{(a)}, \tag{3.8}$$

where X and Y represent amino acid residues, enabling an assessment of electronic correlations at a biologically relevant scale.

3.2.2 Computational Workflow of the dipeptides

A dataset of 400 dipeptides was subjected to geometry optimization using semi-empirical methods (AM1 [45], PM3 [46]), followed by refinement at the B3LYP/6-31G(d) [47] level. MI indices were computed using Kohn-Sham wavefunctions extracted from DFT calculations with the ωB97M-V functional. In analogy to our previous work, SPatial Atom Wise Network (SPAWN) plots were generated for each

dipeptide system to visually interpret these correlations.

To assess the relationship between electronic correlation and molecular stability, AMI values were compared with atomization energies derived from DFT calculations with the ωB97M-V functional.

In the next step, we assessed its correlation with interaction energies using the EDA as provided in Amsterdam Density Functional (ADF) [48] to validate FMI (I_{XY}). This involved splitting dipeptides into two single amino acids by cleaving the peptide bond. Additionally, non-bonded interactions were analyzed using the $1/R^6$ component of the Lennard-Jones potential, with σ and ϵ parameters derived from the OPLS-AA and CHARMM27 force fields.

$$V_{\mathrm{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{3.9}$$

3.2.3 Computational Workflow of the polypeptide

The AMI measurement was further examined on a more biologically significant system, the 10-mer Neh2 domain sequence (QLDEETGEFL). For studying this polypeptide, MD simulations of the Neh2 domain were conducted using GROMACS 2024.4 [49, 50] in explicit water (TIP3P model) with OPLS-AA [51, 30, 52] and CHARMM27 [53, 54, 55] force fields.

Simulations were performed at 310 K and 1 bar pressure and maintained with the v-rescale thermostat [56] with a 0.1 ps coupling time constant, and the Parrinello–Rahman barostat [57] with a 3.22 ps coupling constant. The periodic boundary conditions were applied in all directions. Sodium (Na^+) and chloride (Cl^-) ions were added to neutralize the system and bring the salt concentration to 0.1 M. Energy minimization was achieved through 1000 steepest descent steps, followed by production runs with a timestep of 2 fs. Bond constraints were handled via the LINCS [58] algorithm and long-range electrostatics were computed using the Particle-Mesh

Ewald (PME) [59] method.

The native Beta-turn structure of the Neh2 peptide was obtained from the Protein Data Bank (PDB ID: 2FLU) [60], and its primary (linear) structure was modeled using PyMOL [61]. To capture the folding dynamics, ten independent MD simulations were initiated, each 10 ns in duration. The replicas that exhibited the lowest root mean square deviation (RMSD) from the native Beta-turn structure were extended to 100 ns. Despite extending some simulations up to 500 ns, the native Beta-turn conformation was not achieved, suggesting the complexity of the folding landscape and potential kinetic barriers [62, 63, 64].

Snapshots from the most promising simulation trajectories were extracted at intervals of every 10 ps during the first nanosecond and every 1 ns thereafter, up to 100 ns. Each snapshot was subjected to DFT calculations at the ωB97M-V/6-31G(d) level to compute FMI indices. This measurement provided insight into how electronic correlation evolves throughout the MD simulation.

To further explore the peptide structural conformations during the simulations, clustering analysis was performed using the GROMACS clustering tool. This analysis groups similar structures, identifying representative conformations that capture the diversity of the peptide's structural ensemble. Representative structures from each cluster were selected for DFT computations, where FMI was evaluated using ωB97M-V/ and 6-31G(d). Subsequently, DLPNO-CCSD(T) [65] calculations, known for its efficiency and precision in handling large molecular systems, were performed as a benchmark for energy estimations. The cc-pVDZ basis set and the corresponding cc-pVDZ/C auxiliary basis were used, with default NormalPNO thresholds (TCutPairs = 10^{-4}, TCutPNO = 10^{-7}, TCutMKN = 10^{-3}), along with TightSCF convergence. All ab initio quantum calculations were performed using the ORCA 6.0.1 software package [66].

To visualize the complex network of electronic interactions, SPAWN plots were

adapted to represent interactions at the amino acid residue level, treating each residue as a coarse-grained unit. This approach allowed for a comprehensive visualization of inter-residue electronic correlations, highlighting key interactions that may play a role in the peptide's folding and function.

3.3 Results and discussions

The analysis of 400 dipeptides using AMI revealed distinct electronic correlation patterns in both bonded and non-bonded interactions and confirmed the locality of electronic correlations, reinforcing the nearsightedness of quantum interactions in peptide systems [25]. These findings laid the foundation for further exploring the relationship between AMI and molecular properties in this study. We also extended its application to a larger biomolecular system, the 10-mer Neh2, exploring its potential in understanding protein folding mechanisms and molecular stability.

To establish a quantitative relationship between electronic correlation and molecular stability, the total AMI, obtained by summing all AMI contributions within a dipeptide,

$$\mathrm{AMI}_{\mathrm{tot}} = \sum_{A=1}^{N-1} \sum_{B>A}^{N} I_{AB}, \tag{3.10}$$

where N is the total number of atoms, was compared against the total atomization energy derived from DFT calculations. The analysis demonstrated a strong linear correlation between total AMI and atomization energy ($R^2 = 0.99$). This correlation is depicted in Figure 3.1.

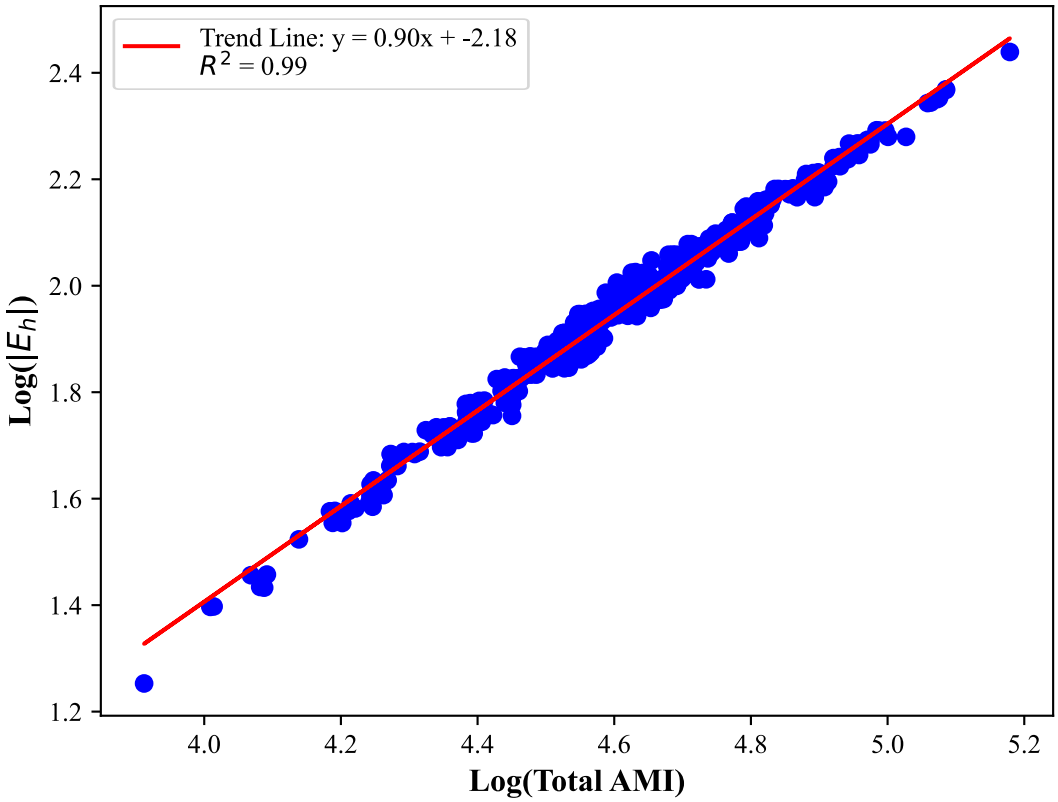


Figure 3.1: Total atomization energy versus compounded AMI for all 400 dipeptide systems

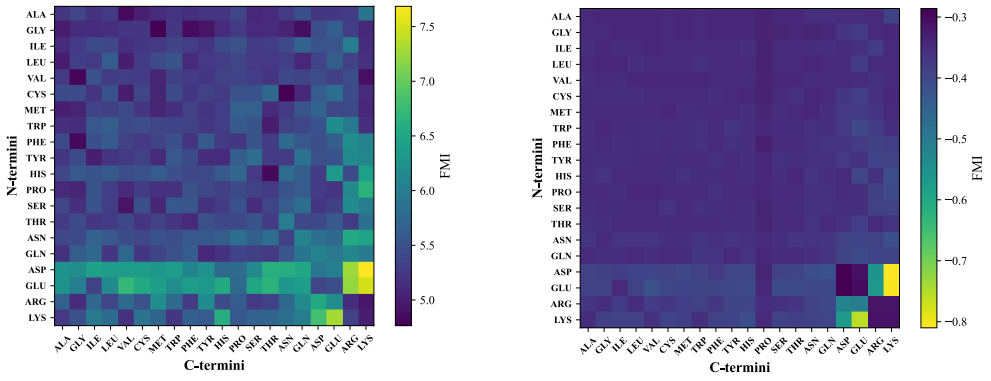
To further investigate inter-amino acid correlations, a coarse-grained extension of AMI, I_{XY} referred to as FMI, was computed using eq. 3.8. Figure 3.2a illustrates the FMI measured for each dipeptide, where dark blue indicates lower correlation and yellow signifies higher correlation. Asymmetry arises because the two ends of the dipeptides (N-termini and C-termini) have different structural configurations, even though the dipeptides consist of the same amino acids.

To understand whether FMI correlates with bonding energy, we examined the bonding interactions between the amino acids within each dipeptide using EDA. Figure 3.2b displays the bonding energy for each dipeptide, where dark blue represents higher bonding energy and yellow indicates lower bonding energy.

The comparison between bonding energy and FMI is further explored in Figure 3.3a, which shows that dipeptides with Coulombic attractions in their side chains (where

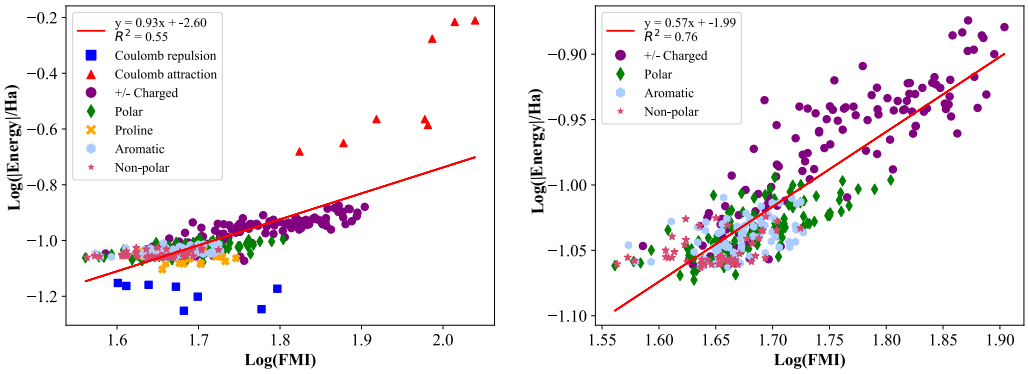
one side chain is positively charged and the other negatively charged) cluster in the upper right corner, indicating a higher level of bonding energy and FMI. In contrast, dipeptides with Coulombic repulsions are located in the lower left corner, showing a lower level of bonding energy and FMI. This observation is also evident in Figure 3.2, where dipeptides clustered in the bottom right corner show varying Coulombic interactions. Those with stronger attractions appear more yellow, indicating higher magnitudes of both FMI and bonding energy, while those with repulsive interactions appear dark blue, reflecting lower values of these quantities.

The correlation between bonding energy and FMI significantly improved when dipeptides with Coulombic interactions are excluded, as seen in Figure 3.3. Furthermore, excluding proline from the data set increased the correlation further. This can be due to the unique structure of proline where its side chain forms a bond with its N-terminus. This can also be observed in Figures 3.2, where the proline column has a distinct shade of blue. Proline's bonding energy is weaker when located at the C-termini, where its nitrogen is engaged in both the peptide bond and its side chain. This was not observed when proline was located at the N-termini, where it forms a peptide bond through its carbon. This distinctive behavior of proline highlights its impact on the overall correlation and bonding interactions within dipeptides.



(a) (b)

Figure 3.2: Visualization of dipeptides' FMI (a) bonding energy (b). Dark blue indicates a lower level of correlation or bonding energy, while yellow signifies a higher degree of correlation or bonding energy.



(a) (b)

Figure 3.3: Correlation between bonding energy and FMI for all 400 dipeptides (a) and after excluding dipeptides with Coulombic attraction/repulsion and proline residues (b), which shows an improvement in correlation.

In the next step, we explored the potential of MI indices to inform the design of classical force fields, which remains a significant challenge to date, even in the presence of emerging machine learning potential. Classical force fields are mathematical models that include various potential energy functions that describe interactions between atoms. One of the core components of classical force fields is the Lennard-Jones potential, which specifically describes the pairwise interactions between non-bonded atoms. Each force field uses specific parameters to describe the Lennard-Jones poten-

tial. The parameters within these force fields need to be carefully tuned to accurately reproduce the behavior of real molecules. However, this tuning process can be intricate due to the complex nature of molecular interactions and the limitations of empirical data. This study explores MI as a potential tool for developing and tuning classical force fields. The hypothesis is that MI could serve as an additional metric for fine-tuning classical force fields, akin to how energy values are used as fitting parameters to adjust the parameters in force field models. The OPLSA [67] and CHARMM [53] are two widely used force fields in molecular simulations. The relationship between AMI and the attractive $1/R^6$ component of the Lennard-Jones potential was analyzed. Only the part of the interaction beyond 4 Å was considered. A notable correlation was observed with the OPLSA force field, while the CHARMM force field exhibits a lesser degree of correlation. Figures 3.4a and 3.4c provide specific examples of this correlation. Further analysis unveils a structural dependency in the correlation between AMI and the Lennard-Jones potential. Dipeptides with more polymeric structures exhibit higher levels of correlation, indicating the locality of AMI.

The Lys-Lys dipeptide 3.4a, which has a more polymeric structure, demonstrates the highest level of correlation with an R^2 value of 0.89. In contrast, the less polymeric Serine-Glycine dipeptide 3.4c shows the lowest level of correlation with an R^2 value of 0.39. Blue dots represent pairwise atomic interaction when both atoms are heavy (carbon, nitrogen, oxygen, and sulfur), while grey dots represent pairwise atomic interaction involving at least one hydrogen atom.

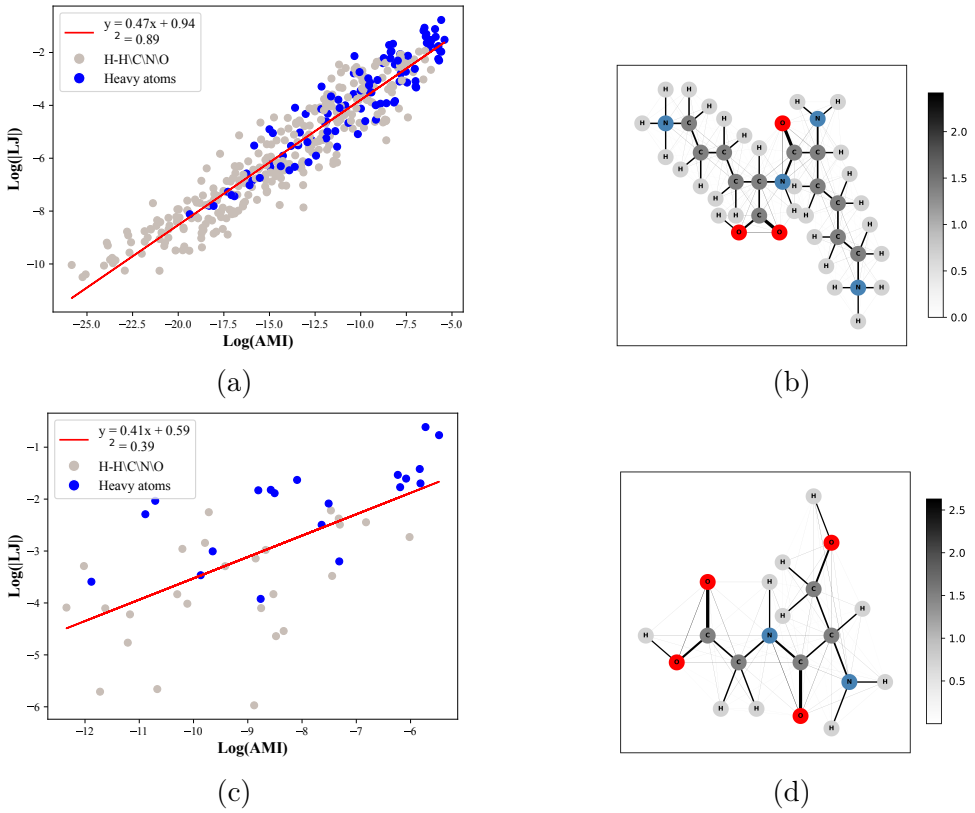
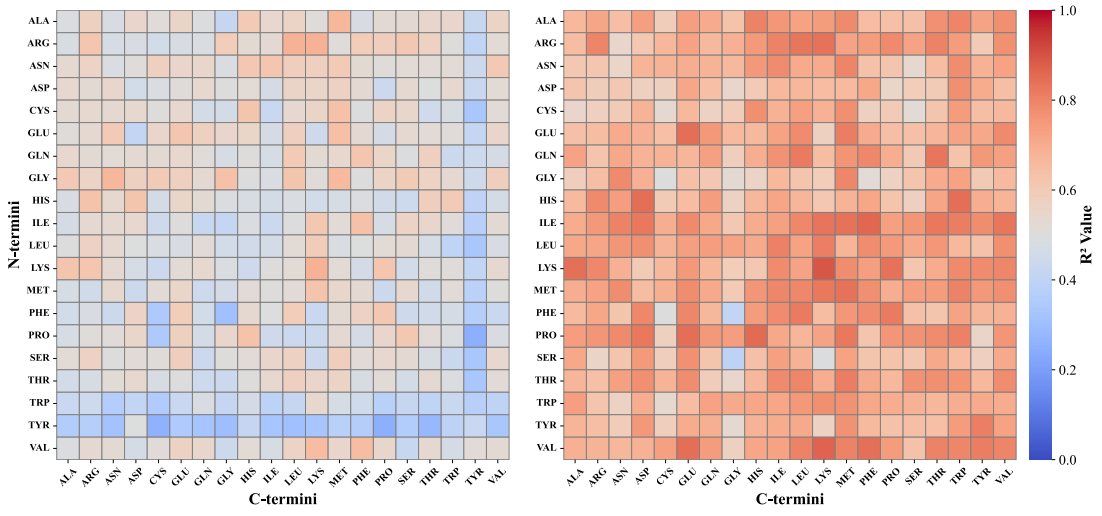


Figure 3.4: Correlation plots between Lennard-Jones energies, derived from OPLS-AA force field, and AMI for Lys-Lys (a) and Ser-Gly (c). Corresponding SPAWN plots are given respectively in (b) and (d).

To systematically compare these correlations, we constructed heatmaps of the R^2 values obtained from the regression analysis between AMI and Lennard-Jones energy for both force fields. Figure 3.5 presents the distribution of R^2 values for CHARMM and OPLS-AA, highlighting the differences in how each force field captures atomic correlations. The results indicate that the OPLS-AA force field exhibits a stronger correlation between AMI and Lennard-Jones energy, with a mean R^2 value of 0.8, whereas CHARMM shows a weaker correlation with a mean R^2 just below 0.6.

This finding suggests that OPLS-AA better reproduces electronic correlation effects as captured by AMI. These results highlight the potential of AMI as a quantum-information metric for evaluating and refining classical force fields, offering a complementary perspective beyond traditional energy-based parameter tuning.



(a) (b)

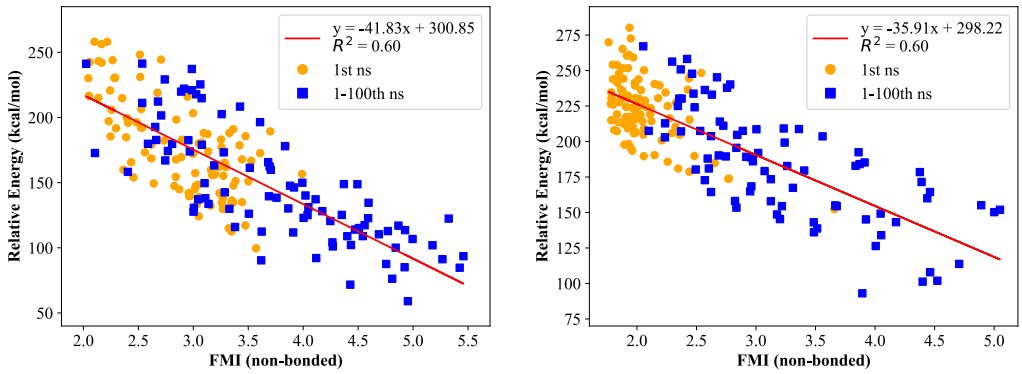
Figure 3.5: Comparison of R^2 value distributions for AMI and Lennard-Jones energy correlations in two force fields: CHARMM (a) and OPLS-AA (b). The heatmaps illustrate the degree of correlation between AMI and the attractive $1/r^6$ component of the Lennard-Jones potential, with higher R^2 values indicating stronger agreement.

Next, we turned our investigation to the evolution of electronic correlations during peptide folding, and conducted MD simulations of the Neh2 peptide using two different force fields, CHARMM and OPLS-AA. The MD started from an unfolded primary structure, and as the simulation progressed, the peptide explored different conformations. At regular intervals, snapshots of these evolving conformations were extracted to track structural changes over time. The energy of each snapshot was then computed using DFT calculations at the ωB97M-V/ and 6-31G(d) level. To establish a baseline for comparison, the reference energy was obtained from the native Beta-turn conformation.

Figure 3.6 presents the energy deviations of these snapshots from the reference Beta-turn structure and correlates them with FMI values for non-bonded interactions, that is, the FMI values between amino acid residues that are not directly connected by covalent bonds. Orange markers represent snapshots extracted every 10 ps within the first nanosecond, capturing the peptide's rapid conformational exploration, which

corresponds to higher energy fluctuations in the early stages. As the simulation progresses, the peptide gradually adopts lower-energy conformations, indicating increased structural stability. Blue markers represent snapshots extracted every 1 ns, extending from 1 to 100 ns, showing a gradual stabilization over time.

The overall trend reveals that as the peptide approaches its folded structure, FMI values for non-bonded interactions increase while energy decreases, reinforcing the role of electronic correlation in stabilizing peptide conformations. This effect is more pronounced in the OPLS-AA force field (Figure 3.6b), which exhibits better convergence toward the reference energy compared to CHARMM (Figure 3.6a). These results suggest that inter-atomic correlations, as quantified by FMI, could serve as a robust descriptor for evaluating force field accuracy in biomolecular simulations.



(a) (b)

Figure 3.6: Energy deviations of Neh2 peptide snapshots relative to the reference Beta-turn structure, correlated with FMI values for non-bonded interactions. The plots compare the energy evolution during MD simulations under two different force fields: CHARMM (a) and OPLS-AA (b). Orange circles represent snapshots extracted at 10 ps intervals within the first nanosecond, capturing initial fluctuations as the peptide explores its conformational space. Blue squares indicate snapshots extracted at 1 ns intervals from 1 to 100 ns, showing a gradual stabilization over time. The simultaneous decrease in energy and increase in FMI values suggest the progressive formation of stable electronic correlations, with the CHARMM force field exhibiting stronger convergence toward lower-energy conformations compared to OPLS-AA.

To identify the metastable conformations, we performed a clustering analysis of MD trajectories, grouping similar structures and selecting representative conformations

from the main clusters. These selected conformations were further analyzed using DLPNO-CCSD(T) and DFT calculations to determine their energies, providing insight into their energetic stability. Then, the FMI values for non-bonded interactions were computed to investigate how inter-atomic correlations evolve with folding and how it correlates with the total energy for each conformation.

Figure 3.7 presents the relative energy of these representative conformations, calculated using DLPNO-CCSD(T) (orange) and DFT (blue), referenced against the native Beta-turn structure. The green line represents the evolution of FMI values for non-bonded interactions as a function of the peptide's end-to-end distance, which serves as a measure of folding compactness.

The results indicate a strong relationship between peptide stability and inter-atomic correlations encoded by our FMI measure. Lower-energy conformations, which are structurally closer to the Beta-turn, as indicated by end-to-end distance as a measure of folding, exhibit higher FMI values, suggesting that stronger inter-atomic correlations contribute to greater structural stability. The progressive increase in FMI values as the peptide transitions from an extended to a folded state reinforces the idea that electronic correlations play a key role in stabilizing the folded structure.

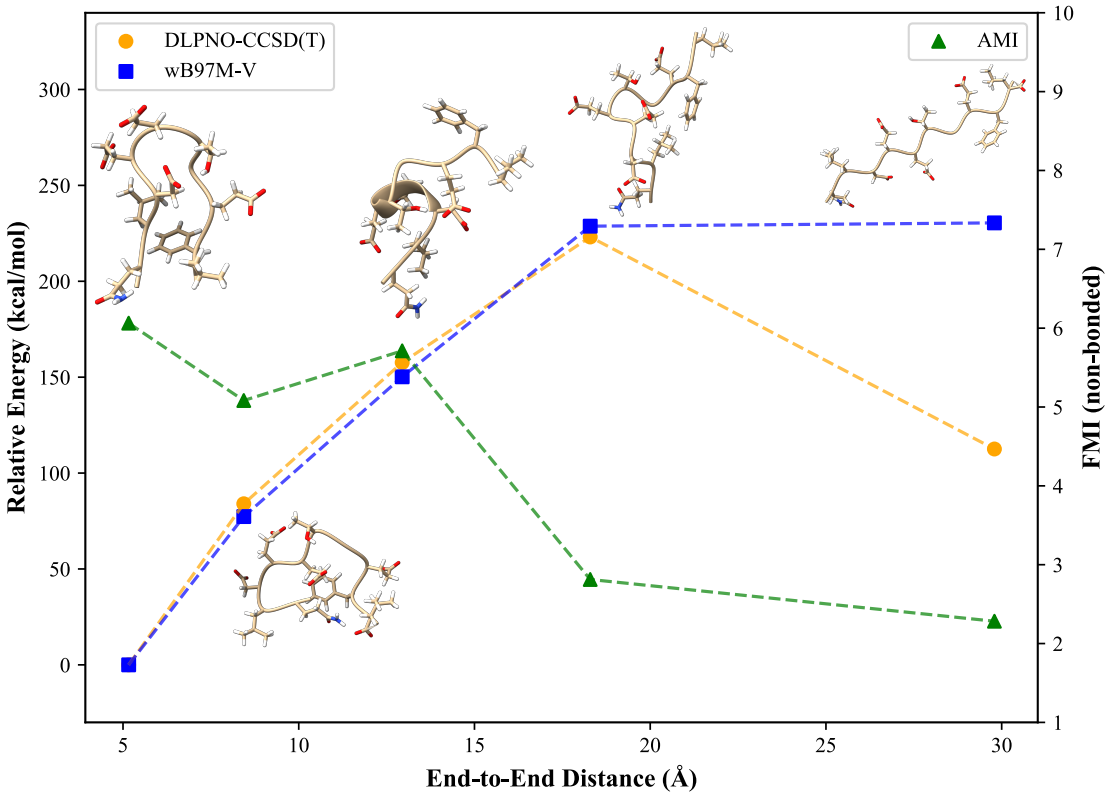


Figure 3.7: DLPNO-CCSD(T) (orange) and DFT (blue) energies for representative peptide conformations identified through clustering analysis of MD trajectories, referenced against the native Beta-turn structure. From left to right, the conformations are beta-turn, cluster 3, cluster 1, cluster 2, and unfolded primary structures. The green line illustrates the evolution of FMI values for non-bonded interactions as a function of the peptide's end-to-end distance, providing insight into its folding behavior. The observed increase in FMI values as energy decreases highlights the role of inter-atomic correlations in stabilizing peptide conformations, with stronger correlations emerging as the peptide approaches its folded state.

Figure 3.9 compares the three-dimensional (3D) structures of the native Beta-turn (a), three metastable clusters (b-d), and the unfolded structure (e), alongside their corresponding SPAWN plots. The 3D structures, visualized using ChimeraX, depict hydrogen bonds as blue dashed lines, providing a conventional representation of stabilizing interactions. However, the SPAWN plots not only mirror these bonds but also offer a more detailed perspective by quantifying their electronic correlation strength through FMI. Unlike traditional hydrogen bond representations that consider all interactions equally, FMI distinguishes between different stabilizing forces

based on their electronic correlation values, approximated by interatomic correlations. For instance, the interaction between Leu 84 and Gln 75 appears as a regular hydrogen bond in the ChimeraX 3D structure, but FMI analysis reveals it as a strong salt bridge with an FMI value of 1.33 nat, highlighting its significant stabilizing role. This enhanced correlation is attributed to the electrostatic attraction between the oppositely charged groups of Gln 75 (NH_3^+) and Leu 84 (COO^-), which would not be evident from a classical hydrogen bond representation alone. Figure 3.8 shows how FMI provides a coarse-grained view of inter-residue correlations, while AMI offers a more fine-grained perspective by resolving the atomic contributions.

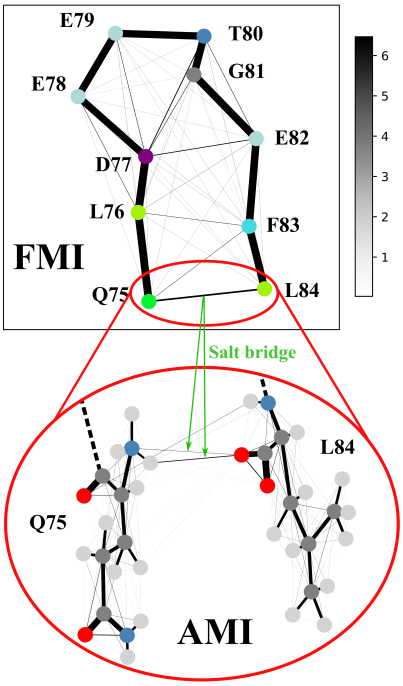


Figure 3.8: Illustration of how FMI emerges as a coarse-grained measure of AMI, capturing inter-residue correlations derived from underlying inter-atomic correlations. In the focused AMI view (red circle), dashed lines represent peptide bonds connecting Gln75 and Leu84 to their neighboring residues. The strong inter-residue FMI between Gln75 and Leu84 is resolved into its dominant AMI components. Green arrows highlight the atomic pairs responsible for the salt bridge interaction between the NH_3^+ group of Gln75 and the COO^- group of Leu84.

Further differentiation is observed in the hydrogen bonds involving Asp 77. The conventional visualization shows that Asp 77 forms four hydrogen bonds, with Thr 80,

Glu 79, Gly 81, and Glu 82, but FMI analysis reveals that not all of these contribute equally to structural stabilization. The interactions with Thr 80 and Glu 82 exhibit higher FMI values (0.6 nat), indicating stronger electronic correlations, whereas the interactions with Glu 79 and Gly 81 are weaker (0.3 nat). The SPAWN plots provide a clearer picture of these variations, where thicker non-covalent connections correspond to stronger FMI values, offering a level of detail beyond classical structural representations.

Table 3.1: Summary of key hydrogen bond interactions and corresponding FMI values for metastable clusters.

Conformations	**H-Bond Residue Pair**	**Distance ≈ (Å)**	**FMI ≈(nat)**
Cluster 1 (Fig. 3.9b)	Gln75–Glu79	2.89	0.80
	Leu76–Glu79	2.95	0.39
	Glu79–Leu76	2.78	0.41
	Thr80–Leu76	3.04	0.25
	Thr80–Glu82	2.81	0.79
	Gly81–Asp77	3.00	0.21
	Glu82–Asp77	2.72	0.80
Cluster 2 (Fig. 3.9c)	Glu78–Gly81	3.12	0.17
Cluster 3 (Fig. 3.9d)	Gln75–Glu79	2.58	1.14
	Gln75–Glu78	2.79	0.48
	Gln75–Leu84	2.78	0.70

The metastable clusters sampled from MD simulations further demonstrate the role of FMI in tracking structural transitions. Table 3.1 presents the donor–acceptor residue pairs involved in hydrogen bonding for the three main metastable clusters. Cluster 1 displays a helix-like structure stabilized by seven hydrogen bonds. Cluster 2 has only one weak hydrogen bond. Cluster 3 shows a strong salt bridge with Glu79 (FMI ≈ 1.14 nat), indicating a highly correlated electrostatic interaction.

Although Cluster 1 features some of the highest individual FMI values, Cluster 3 ap-

pears to be more energetically favorable overall, as shown in Figure 3.7. This observation reinforces the complementary role of FMI in assessing molecular interactions, capturing stabilizing electronic correlations that may not always directly translate into lower total energy.

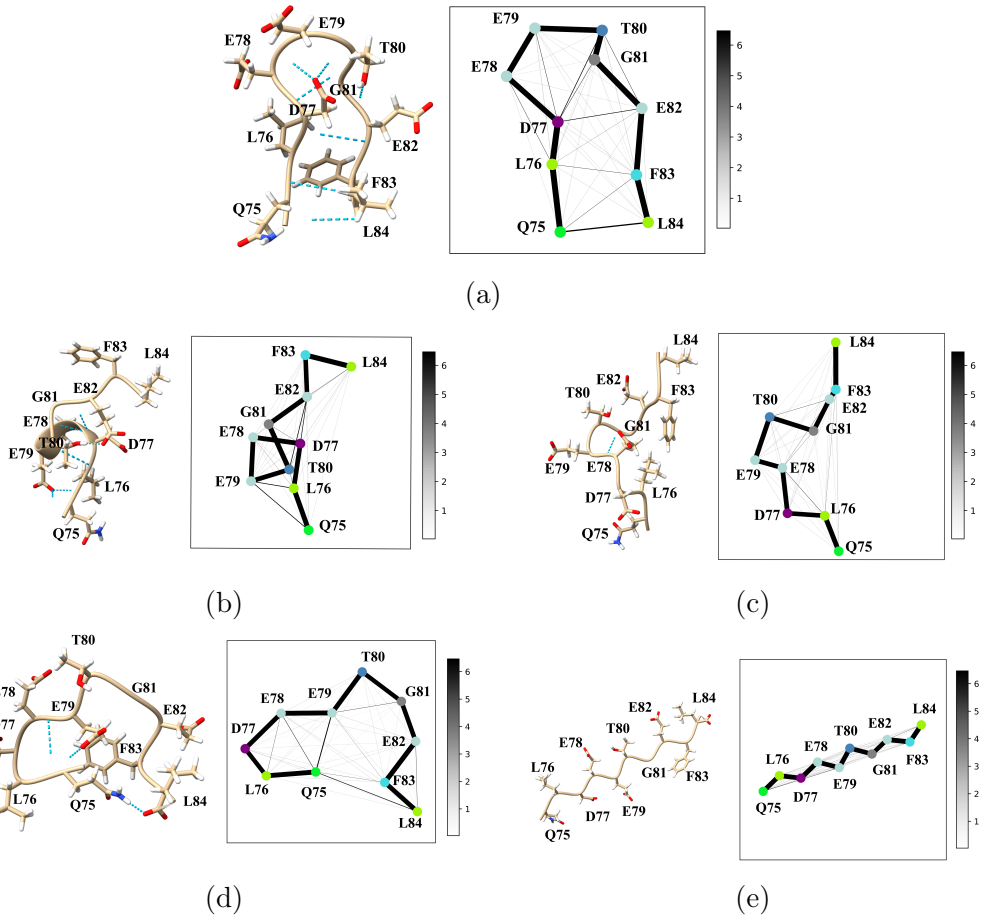


Figure 3.9: Comparison of peptide metastable conformations and their corresponding FMI SPAWN plots. (a) Native Beta-turn, (b) Cluster 1, (c) Cluster 2, (d) Cluster 3, and (e) Unfolded structure. The 3D structures, visualized using ChimeraX, depict backbone atoms as ribbons, with only side chains explicitly shown for clarity. Hydrogen bonds are highlighted with blue dashed lines. The SPAWN plots provide a quantitative view of electronic correlations, where each bead represents an amino acid. The thickest lines correspond to covalent peptide bonds ∼6 nat in the grayscale gradient), while thinner gray lines indicate non-covalent interactions.

Instead, FMI provides a deeper understanding of the interaction landscape by identifying key stabilizing interactions with their associated strengths. The agreement

between FMI values and classical interaction descriptors, such as salt bridges and hydrogen bonds, suggests that integrating FMI into molecular modeling could enhance force field accuracy by offering a more refined perspective on stabilizing factors beyond classical descriptors.

The unfolded structure (e) lacks defined hydrogen bonds, and its SPAWN plot shows weak FMI connectivity, reinforcing the loss of electronic coherence as structural disorder increases. Unlike the structured Beta-turn and metastable clusters, the FMI values in this conformation are low, indicating that electronic interactions weaken as the peptide adopts a disordered, flexible state. The gradual increase in FMI values as the peptide transitions from an unfolded to a folded state highlights the role of electronic correlations in guiding folding pathways. The agreement between FMI and key stabilizing forces, such as salt bridges and varying hydrogen bond strengths, demonstrates its potential for improving force field accuracy in molecular simulations.

We also explored a simplified classical model in which all atom pairs were categorized according to their atom types (C−C, C−N, C−O, C−S, C−H, O−N, ...,H−H). Building on the trends observed in Figure 2.7, where the relationship between log(AMI) and interatomic distance was analyzed for all dipeptides, we repeated this analysis separately for each atom-type pair using the same dataset. This enabled us to characterize the distance dependence of log(AMI) for each pair type, leading to the construction of pair-specific parametric models of the form

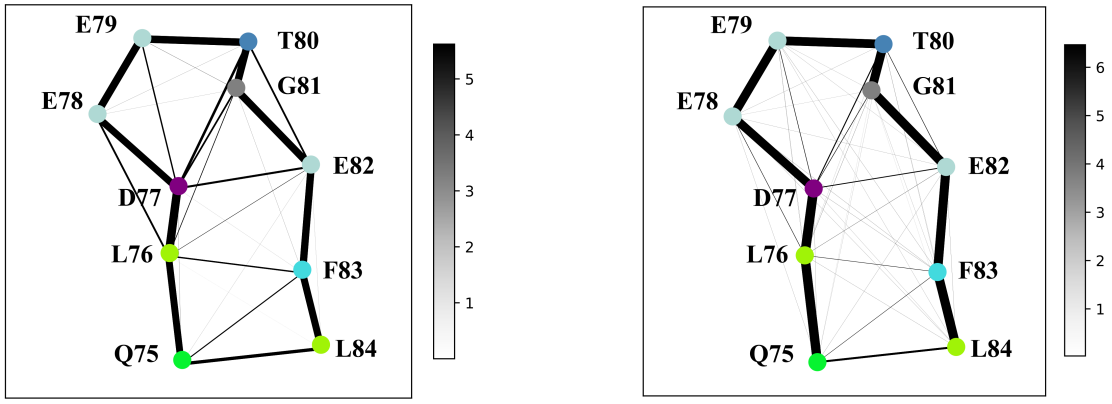
$$\mathrm{AMI}(R) = e^{\beta} \cdot e^{\alpha R}, \tag{3.11}$$

where α and β are fitted parameters corresponding to the slope and intercept of the linear model for each atom-type pair, and R denotes the interatomic distance. These parameters were compiled into lookup tables and subsequently used to estimate AMI values directly from interatomic distances. In this way, the model replaces expensive quantum mechanical calculations with a simple distance-based evaluation, enabling

efficient reconstruction of AMI and FMI values for larger molecular systems.

Figure 3.10 compares FMI values obtained from the classical model with those computed at the DFT ωB97M-V level. The classical model captures the overall structure of the interaction SPAWN plot, including the dominant covalent bonds and main non-covalent interactions.

Discrepancies are more noticeable for weaker interactions, where the distance-based model does not fully account for the influence of the chemical environment and many-body effects. These differences highlight the limitations of a purely pairwise representation. Despite these limitations, the results demonstrate that the classical model provides a computationally efficient approximation that retains the essential FMI values represented in SPAWN plots, supporting its potential use in large-scale biomolecular applications.



(a) (b)

Figure 3.10: Comparison of FMI SPAWN networks for the Neh2 native conformation obtained from the classical distance-based model (a) and DFT ωB97M-V calculations (b).

3.4 Conclusion

This study highlights the capability of AMI and FMI as a powerful tool for analyzing inter-atomic interactions in biomolecules. By extending AMI concepts from atomic-

level correlations to amino acid fragments, we demonstrated that FMI provides a quantitative measure of the role of inter-atomic correlations in stabilizing peptide structures. Our findings reinforce the nearsightedness of quantum interactions in peptide structures and establish a clear correlation between FMI and molecular stability.

Through an extensive analysis of 400 dipeptides, we identified that FMI correlates well with conventional energy descriptors such as atomization energy and bonding energy from EDA, capturing electronic interactions beyond classical descriptors such as hydrogen bonding and van der Waals interactions. Our comparison of CHARMM and OPLS-AA force fields highlights FMI's utility in evaluating force field accuracy, suggesting that FMI could serve as a complementary parameter in force field refinement.

This relationship was further validated in a more complex biomolecular system, the 10-mer Neh2 peptide, where FMI provided insight into the electronic correlation dynamics during peptide folding. Our results showed that FMI can differentiate between folded and unfolded peptide states. By tracking MI changes over time, we demonstrated that electronic interactions play a crucial role in stabilizing the folded structure, with FMI values increasing as the peptide transitions toward a stable Beta-turn conformation.

Future research will focus on integrating FMI with machine learning models to develop predictive tools for protein-ligand interactions, further enhancing its applications in computational drug design and biomolecular engineering. AMI and FMI also hold promise for quantum embedding methods. By identifying electronically correlated atoms or fragments, these measures can guide the definition of chemically meaningful active spaces or embedding regions, thereby enhancing both the accuracy and efficiency of multiscale approaches. These findings pave the way for a quantum-information approach to studying biomolecular interactions, providing a

more accurate and detailed understanding of inter-atomic correlations in biological systems.

3.5 Acknowledgements

S.D.B acknowledges the NSERC Canada Research Chair and Discovery Grant program for financial support. M.J.M gratefully acknowledges a MITACS Globalink grant to allow for an extensive research stay at the Nicolaus Copernicus University in Toruń. Funded/Co-funded by the European Union (ERC, DRESSED-pCCD, 101077420). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. P.T. is grateful for an OPUS 17 research grant of the National Science Centre, Poland (no. 2019/33/B/ST4/02114).

Bibliography

[1] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, Phys. Rev. Lett. 69 (1992) 2863–2866.

[2] S. R. White, Density-Matrix Algorithms for Quantum Renormalization Groups, Phys. Rev. B 48 (1993) 10345–10356.

[3] S. R. White, R. L. Martin, Ab Initio Quantum Chemistry Using the Density Matrix Renormalization Group, J. Chem. Phys. 110 (9) (1999) 4127–4130.

[4] O. Legeza, R. M. Noack, J. Sólyom, L. Tincani, Applications of Quantum Information in the Density-Matrix Renormalization Group, in: Computational Many-Particle Physics, Vol. 739 of Lect. Notes Phys., Springer, Berlin/Heidelerg, 2008, pp. 653–664.

[5] G. K.-L. Chan, S. Sharma, The Density Matrix Renormalization Group in Quantum Chemistry, Annu. Rev. Phys. Chem. 62 (1) (2011) 465–481.

[6] A. Baiardi, M. Reiher, The Density Matrix Renormalization Group in Chemistry and Molecular Physics: Recent Developments and New Challenges, J. Chem. Phys. 152 (4) (2020) 040903.

[7] J. Čížek, On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods, J. Chem. Phys. 45 (11) (1966) 4256–4266.

[8] J. Cizek, J. Paldus, Correlation Problems in Atomic and Molecular Systems III. Rederivation of the Coupled-Pair Many-Electron Theory Using the Traditional Quantum Chemical Methods, Int. J. Quantum Chem. 5 (1971) 359–379.

[9] R. J. Bartlett, M. Musiał, Coupled-Cluster Theory in Quantum Chemistry, Rev. Mod. Phys. 79 (1) (2007) 291–352.

[10] P. Hohenberg, W. Kohn, Inhomogeneous Electron Gas, Phys. Rev. 136 (3B) (1964) B864.

[11] W. Kohn, L. J. Sham, Self-Consistent Equations Including Exchange and Correlation Effects, Phys. Rev. 140 (4A) (1965) A1133.

[12] K. Ramachandran, G. Deepa, K. Namboori, Computational Chemistry and Molecular Modeling: Principles and Applications, Springer Science & Business Media, 2008.

[13] A. S. Christensen, T. Kubar, Q. Cui, M. Elstner, Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications, Chem. Rev. 116 (9) (2016) 5301–5337.

[14] E. D. Glendening, C. R. Landis, F. Weinhold, Natural Bond Orbital Methods, WIREs Comput. Mol. Sci. 2 (1) (2012) 1–42.

[15] R. F. W. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, 1990.

[16] Ö. Legeza, J. Sólyom, Optimizing the Density-Matrix Renormalization Group Method Using Quantum Information Entropy, Phys. Rev. B 68 (2003) 195116.

[17] K. Boguslawski, P. Tecmer, G. Barcza, Ö. Legeza, M. Reiher, Orbital Entanglement in Bond-Formation Processes, J. Chem. Theory Comput. 9 (7) (2013) 2959–2973.

[18] M. Mottet, P. Tecmer, K. Boguslawski, O. Legeza, M. Reiher, Quantum Entanglement in Carbon-Carbon, Carbon-Phosphorus and Silicon-Silicon Bonds, Phys. Chem. Chem. Phys. 16 (2014) 8872–8880.

[19] C. Duperrouzel, P. Tecmer, K. Boguslawski, G. Barcza, O. Legeza, P. W. Ayers, A Quantum Informational Approach for Dissecting Chemical Reactions, Chem. Phys. Lett. 621 (2015) 160–164.

[20] Y. Zhao, K. Boguslawski, P. Tecmer, C. Duperrouzel, G. Barcza, Ö. Legeza, P. W. Ayers, Dissecting the Bond-Formation Process of D 10-Metal–ethene Complexes With Multireference Approaches, Theor. Chem. Acc. 134 (10) (2015) 120.

[21] J.-X. Zhang, F. K. Sheong, Z. Lin, Unravelling Chemical Interactions With Principal Interacting Orbital Analysis, Chem. Eur. J. 24 (38) (2018) 9639–9650.

[22] A. Leszczyk, T. Dome, P. Tecmer, D. Kędziera, K. Boguslawski, Resolving the Π-Assisted U–N Σ F-Bond Formation Using Quantum Information Theory, Phys. Chem. Chem. Phys. 24 (35) (2022) 21296–21307.

[23] R. Chakraborty, K. Boguslawski, P. Tecmer, Static Embedding With Pair Coupled Cluster Doubles Based Methods, Phys. Chem. Chem. Phys. 25 (2023) 25377–25388.

[24] M. Gałyńska, M. M. F. de Moraes, P. Tecmer, K. Boguslawski, Delving Into the Catalytic Mechanism of Molybdenum Cofactors: A Novel Coupled Cluster Study, Phys. Chem. Chem. Phys. 26 (27) (2024) 18918–18929.

[25] M. J. Moghadam, K. Boguslawski, R. Doucet, Ö. Legeza, P. Tecmer, S. De Baerdemacker, Correlation in Dipeptide Systems Form Quantum Information Theory Perspectives, chemrxiv. (2024).

[26] F. M. Bickelhaupt, E. J. Baerends, Kohn-Sham Density Functional Theory: Predicting and Understanding Chemistry, John Wiley & Sons, Ltd, 2000, pp. 1–86.

[27] M. v. Hopffgarten, G. Frenking, Energy Decomposition Analysis, WIREs Comput. Mol. Sci. 2 (1) (2012) 43–62.

[28] J. E. Jones, On the Determination of Molecular Fields.—II. From the Equation of State of a Gas, Proc. R. Soc. Lond. A 106 (738) (1924) 463–477.

[29] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. a. Swaminathan, M. Karplus, CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations, J. Comput. Chem. 4 (2) (1983) 187–217.

[30] W. L. Jørgensen, D. S. Maxwell, J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, J. Am. Chem. Soc. 118 (45) (1996) 11225–11236.

[31] A. K. Dunker, P. Romero, Z. Obradovic, E. C. Garner, C. J. Brown, Intrinsic Protein Disorder in Complete Genomes, Genome Inf. 11 (2000) 161–171.

[32] A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, et al., Intrinsically Disordered Protein, J. Mol. Graph. Model. 19 (1) (2001) 26–59.

[33] A. K. Dunker, Z. Obradovic, The Protein Trinity—linking Function and Disorder, Nat. Biotechnol. 19 (9) (2001) 805–806.

[34] K. Itoh, T. Chiba, S. Takahashi, T. Ishii, K. Igarashi, Y. Katoh, T. Oyake, N. Hayashi, K. Satoh, I. Hatayama, et al., An Nrf2/Small Maf Heterodimer Mediates the Induction of Phase II Detoxifying Enzyme Genes Through Antioxidant Response Elements, Biochem. Biophys. Res. Commun. 236 (2) (1997) 313–322.

[35] K. Itoh, N. Wakabayashi, Y. Katoh, T. Ishii, K. Igarashi, J. D. Engel, M. Yamamoto, Keap1 Represses Nuclear Activation of Antioxidant Responsive Elements by Nrf2 Through Binding to the Amino-Terminal Neh2 Domain, Genes Dev. 13 (1) (1999) 76–86.

[36] N. Mardirossian, M. Head-Gordon, ΩB97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-Gga Density Functional With VV10 Nonlocal Correlation, J. Phys. Chem. 144 (21) (2016) 214110.

[37] R. Ditchfield, W. J. Hehre, J. A. Pople, Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules, J. Chem. Phys. 54 (2) (1971) 724–728.

[38] J. Rissler, R. M. Noack, S. R. White, Measuring Orbital Interaction Using Quantum Information Theory, Chem. Phys. 323 (2006) 519–531.

[39] K. Boguslawski, P. Tecmer, O. Legeza, M. Reiher, Entanglement Measures for Single-and Multireference Correlation Effects, J. Phys. Chem. Lett. 3 (21) (2012) 3129–3135.

[40] G. Barcza, R. M. Noack, J. Sólyom, Ö. Legeza, Entanglement Patterns and Generalized Correlation Functions in Quantum Many-Body Systems, Phys. Rev. B 92 (12) (2015) 125140.

[41] K. Boguslawski, P. Tecmer, Orbital Entanglement in Quantum Chemistry, Int. J. Quantum Chem. 115 (19) (2015) 1289–1295.

[42] K. Boguslawski, P. Tecmer, Erratum: Orbital Entanglement in Quantum Chemistry, Int. J. Quantum Chem. 117 (24) (2017) e25455.

[43] K. Boguslawski, P. Tecmer, Ö. Legeza, Analysis of Two-Orbital Correlations in Wave Functions Restricted to Electron-Pair States, Phys. Rev. B 94 (15) (2016) 155126.

[44] A. Nowak, Ö. Legeza, K. Boguslawski, Orbital Entanglement and Correlation From pCCD-tailored Coupled Cluster Wave Functions, J. Chem. Phys. 154 (8) (2021) 084111.

[45] M. J. Dewar, E. G. Zoebisch, E. F. Healy, J. J. Stewart, Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model, J. Am. Chem. Soc. 107 (13) (1985) 3902–3909.

[46] J. J. Stewart, Optimization of Parameters for Semiempirical Methods I. Method, J. Comput. Chem. 10 (2) (1989) 209–220.

[47] A. D. Becke, Density-Functional Thermochemistry. III. The Role of Exact Exchange, J. Chem. Phys. 98 (7) (1993) 5648–5652.

[48] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, T. Ziegler, Chemistry With ADF, J. Comput. Chem. 22 (9) (2001) 931–967.

[49] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers, Softw. X 1 (2015) 19–25.

[50] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, GROMACS: Fast, Flexible, and Free, J. Comput. Chem. 26 (16) (2005) 1701–1718.

[51] W. L. Jørgensen, J. Tirado-Rives, The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin, J. Am. Chem. Soc. 110 (6) (1988) 1657–1666.

[52] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison With Accurate Quantum Chemical Calculations on Peptides, J. Phys. Chem. B 105 (28) (2001) 6474–6487.

[53] A. D. Mackerell Jr, M. Feig, C. L. Brooks III, Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations, J. Comput. Chem. 25 (11) (2004) 1400–1415.

[54] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, E. Lindahl, Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects From Correction Maps, Virtual Interaction Sites, and Water Models, J. Chem. Theory Comput. 6 (2) (2010) 459–466.

[55] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, et al., All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins, J. Phys. Chem. B 102 (18) (1998) 3586–3616.

[56] G. Bussi, D. Donadio, M. Parrinello, Canonical Sampling Through Velocity Rescaling, J. Chem. Phys. 126 (1) (2007) 014101.

[57] M. Parrinello, A. Rahman, Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method, J. Appl. Phys. 52 (12) (1981) 7182–7190.

[58] B. Hess, H. Bekker, H. J. Berendsen, J. G. Fraaije, LINCS: A Linear Constraint Solver for Molecular Simulations, J. Comput. Chem. 18 (12) (1997) 1463–1472.

[59] T. Darden, D. York, L. Pedersen, et al., Particle Mesh Ewald: An N Log (N) Method for Ewald Sums in Large Systems, J. Chem. Phys. 98 (1993) 10089–10089.

[60] S.-C. Lo, X. Li, M. T. Henzl, L. J. Beamer, M. Hannink, Structure of the Keap1: Nrf2 Interface Provides Mechanistic Insight Into Nrf2 Signaling, EMBO J. 25 (15) (2006) 3605–3617.

[61] L. Schrödinger, W. DeLano, PyMOL www.pymol.org 2026-04-28.

[62] E. A. Cino, W.-Y. Choy, M. Karttunen, Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations, J. Chem. Theory Comput. 8 (8) (2012) 2725–2740.

[63] T. N. Do, W.-Y. Choy, M. Karttunen, Accelerating the Conformational Sampling of Intrinsically Disordered Proteins, J. Chem. Theory Comput. 10 (11) (2014) 5081–5094.

[64] M. Karttunen, W.-Y. Choy, E. A. Cino, Prediction of Binding Energy of Keap1 Interaction Motifs in the Nrf2 Antioxidant Pathway and Design of Potential High-Affinity Peptides, J. Phys. Chem. B 122 (22) (2018) 5851–5859.

[65] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese, Sparse Maps—A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II.

Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory, J. Chem. Phys. 144 (2) (2016) 024109.

[66] F. Neese, The ORCA Program System, WIREs Comput. Mol. Sci. 2 (1) (2012) 73–78.

[67] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, J. Am. Chem. Soc. 118 (45) (1996) 11225–11236.

Chapter 4

Mapping Electronic Correlations in Proteins via Cut-Wise Reconstruction: A Case Study on Insulin

Mostafa Javaheri Moghadam[1], Rebecca Mulder[2], Stijn De Baerdemacker[1,3]

[1]Department of Chemistry, University of New Brunswick, Fredericton, NB, Canada

[2]Department of Biology, University of New Brunswick, Fredericton, NB, Canada

[3]Department of Mathematics and Statistics, University of New Brunswick, Fredericton, NB, Canada.

Abstract

This study quantifies electronic correlations in the insulin protein using mutual information (MI) to evaluate interatomic and inter-residue interaction strength. More specifically, we introduced and analyzed a cut-wise approach rooted in the principle of quantum locality, assuming that electronic correlations decay rapidly with distance. Through localized DFT calculations on 49 overlapping spherical cuts, we reconstruct global MI matrices at both the atomic and residue levels. This method captures biochemically meaningful features while closely matching results from full-protein DFT calculations, demonstrating the scalability of quantum correlation analysis for large biomolecular systems. The framework sets the stage for future applications in protein–ligand interaction modeling, pharmacophore development, and quantum-enhanced drug discovery.

4.1 Introduction

Understanding the role of electronic interactions in biomolecular systems is fundamental to uncovering the physical principles behind protein folding, enzymatic activity, and molecular recognition [1]. These interactions shape the structural stability and functional specificity of biological macromolecules, driving a wide range of cellular processes [2, 3]. However, the inherently quantum mechanical nature of electron correlation poses a significant challenge for computational modeling, especially when extending these insights to large systems such as proteins [4].

Traditional modeling approaches, including classical force fields and semi-empirical methods, provide scalable solutions but often neglect key quantum effects that are critical for describing subtle intra- and intermolecular interactions [5]. This limits their reliability in applications such as drug design and protein modeling, where the accuracy of interaction profiles can determine the success of lead optimization [6, 7, 8].

In contrast, ab initio quantum chemistry methods offer a first-principles framework for resolving electronic interactions with high accuracy [9]. Yet, their exponential computational scaling constrains their use to small molecules or molecular fragments [10]. Bridging this gap between accuracy and scalability is an ongoing pursuit in theoretical chemistry.

To address this limitation, we adopt a strategy inspired by quantum information theory, MI, which has been adopted in quantum chemistry to quantify electronic correlations between orbitals and molecular fragments [11]. MI has been used extensively in the context of molecular orbital theory and tensor network approaches like the density matrix renormalization group (DMRG) [12, 13, 14] to reveal nontrivial entanglement patterns that govern molecular reactivity and bonding.

Expanding on this concept, we recently introduced atomic mutual information (AMI) [15], a framework that quantifies the shared electronic information between atoms without assuming any specific type of chemical interaction. AMI quantifies electronic correlations between atoms and offers a high-resolution perspective on the connectivity and interaction strength between atoms across a molecular system.

As a proof of principle, in this work, we apply the AMI framework to insulin, a biologically vital peptide hormone composed of 51 amino acids [16]. Insulin plays a central role in regulating blood glucose levels and is a well-characterized benchmark in structural biology [17]. Importantly, its relatively small size makes it an ideal candidate for testing our fragment-based reconstruction approach. It is large enough to exhibit complex tertiary structure and a diverse set of chemical interactions, yet still tractable for density functional theory (DFT) calculations [18, 19].

Despite the limitations of applying DFT directly to full proteins, we leverage the principle of locality in quantum correlations, supported by the area law of entanglement entropy, to decompose the insulin molecule into overlapping fragments. This strategy is conceptually related to methods such as the divide-and-conquer approach, where

the total electron density is reconstructed from subsystem contributions [20, 21, 22]. Using the same concept, DFT calculations are performed on each fragment, and the results are assembled to reconstruct a global AMI-based correlation map of the full protein.

This study demonstrates the feasibility of applying AMI to realistic biomolecular systems and highlights its potential as a quantum-informed tool for analyzing protein structure, interaction networks, and chemical environments with unprecedented detail. The insights gained here lay the foundation for future applications in protein–ligand analysis and rational drug design.

4.2 Methodology

Our approach builds upon a series of prior studies exploring quantum entanglement and electronic correlations in small biomolecular systems such as dipeptides [23, 24, 25, 15]. In this work, we scale the framework to a full protein, employing MI and its coarse-grained form to quantify the correlation landscape within insulin.

4.2.1 Mutual Information from Reduced Density Matrices

At the core of our method is the quantification of electronic correlations using von Neumann entropy. For a bipartite quantum system described by a pure state with the Hilbert space of the subsystem and environment spanned by the states $|n\rangle$ and $|e\rangle$ respectively,

$$|\Psi\rangle = \sum_{n,e} \Psi_{n,e} |n\rangle \otimes |e\rangle, \tag{4.1}$$

we obtain the reduced density matrix (RDM) of the subsystem by tracing out the environment:

$$\rho = \mathrm{Tr}_{|e\rangle} \left(|\Psi\rangle\langle\Psi| \right). \tag{4.2}$$

From the eigenvalues ω_α of this RDM, the entanglement entropy of a single orbital is calculated as:

$$s^{(1)}i = -\sum_{\alpha=1}^{4} \omega_{\alpha,i} \ln \omega_{\alpha,i}, \tag{4.3}$$

while the joint entropy of a pair of orbitals with 4 possible electron configurations $\{ |-\rangle, |\uparrow\rangle, |\downarrow\rangle, |\updownarrow\rangle \}$ is defined as:

$$s^{(2)}ij = -\sum_{\alpha=1}^{16} \omega_{\alpha,ij} \ln \omega_{\alpha,ij}. \tag{4.4}$$

These entropies are then used to compute the MI:

$$I_{ij} = s_{ij}^{(2)} - s_i^{(1)} - s_j^{(1)}, \tag{4.5}$$

which captures both quantum and classical correlations between orbitals i and j. This provides a resolution-independent means of identifying electronically significant interactions in the molecular system.

4.2.2 From Orbital to Atomic and Residue-Level Correlations

To interpret MI in a chemically and biologically meaningful way, we coarse-grain the orbital-based values to the atomic and residue levels. The AMI is defined as the sum over all orbital pairs belonging to atoms A and B:

$$I_{AB} \approx \sum_{i \in A} \sum_{j \in B} I_{ij}. \tag{4.6}$$

Further coarse-graining yields fragment-wise mutual information (FMI), which aggregates atomic contributions over residues X and Y:

$$I_{XY} \approx \sum_{A \in X} \sum_{B \in Y} I_{AB}. \tag{4.7}$$

This hierarchy, from orbital to atomic to residue scale, enables the analysis of electronic structure across multiple levels of resolution, from covalent bonds to long-range intra-protein interactions.

4.2.3 System Preparation and Fragmentation

The 3D structure of insulin (PDB ID: 3I40) [16] was obtained from the Protein Data Bank. The structure was protonated and energy minimized using the steepest descent algorithm implemented in GROMACS [26, 27] with the OPLS-AA force field [28, 29] and TIP3P water model. To manage the system's size (782 atoms), we divided the protein into 49 overlapping spherical fragments, each centered on an α-carbon and extending to a radius of 5.0 Å. This ensured both locality and overlap, enabling recombination of quantum information across the full protein.

Each fragment was subjected to DFT calculations using ORCA 6.0.1 [30] with the ωB97M-V functional [31] and 6-31G(d) basis set [32]. AMIs were computed for each fragment, which were then stitched together to construct AMI and FMI matrices for the full system. Overlapping contributions from overlapping regions were averaged to ensure consistency.

To validate the accuracy of our fragment-based reconstruction approach, we conducted a full DFT calculation on the entire insulin protein without fragmenting the structure. Using the same functional and basis set, we extracted the complete wavefunction and computed the MI matrix from the full-system RDMs. This full-protein MI map served as a benchmark to evaluate the consistency and fidelity of the stitched correlation data assembled from the overlapping structural cuts.

4.3 Results and discussions

The three-dimensional structure of insulin, along with the defined structural cuts, provides the foundation for the cut-wise quantum mechanical analysis used to evaluate electronic correlations across the whole protein. Figure 4.1 shows a ribbon representation of the insulin structure, highlighting its two-chain (A and B) and overall tertiary fold. The figure illustrates how the protein was segmented into overlapping spherical fragments centered on α-carbon atoms, with two representative cuts marked. A magnified view of Cut 1 shows the neighboring residue environment captured within a 5.0 Å radius. Each cut encapsulates a chemically meaningful region while maintaining sufficient overlap to reconstruct global electronic correlations. As each individual cut represents the 5.0 Å cut-off distance from its central α-carbon, it may result in fragments or residues that are not connected through peptide bonds in that individual cut. For instance, AlaB30 is disconnected from the Chain A residues in Cut 1 (Figure 4.1). Although counterintuitive, including disconnected fragments helps stabilize the overlapping contributions. The corresponding sequence in Figure 4.1 (b) maps Chain A (21 residues) and Chain B (30 residues), providing a direct reference for fragment positioning.

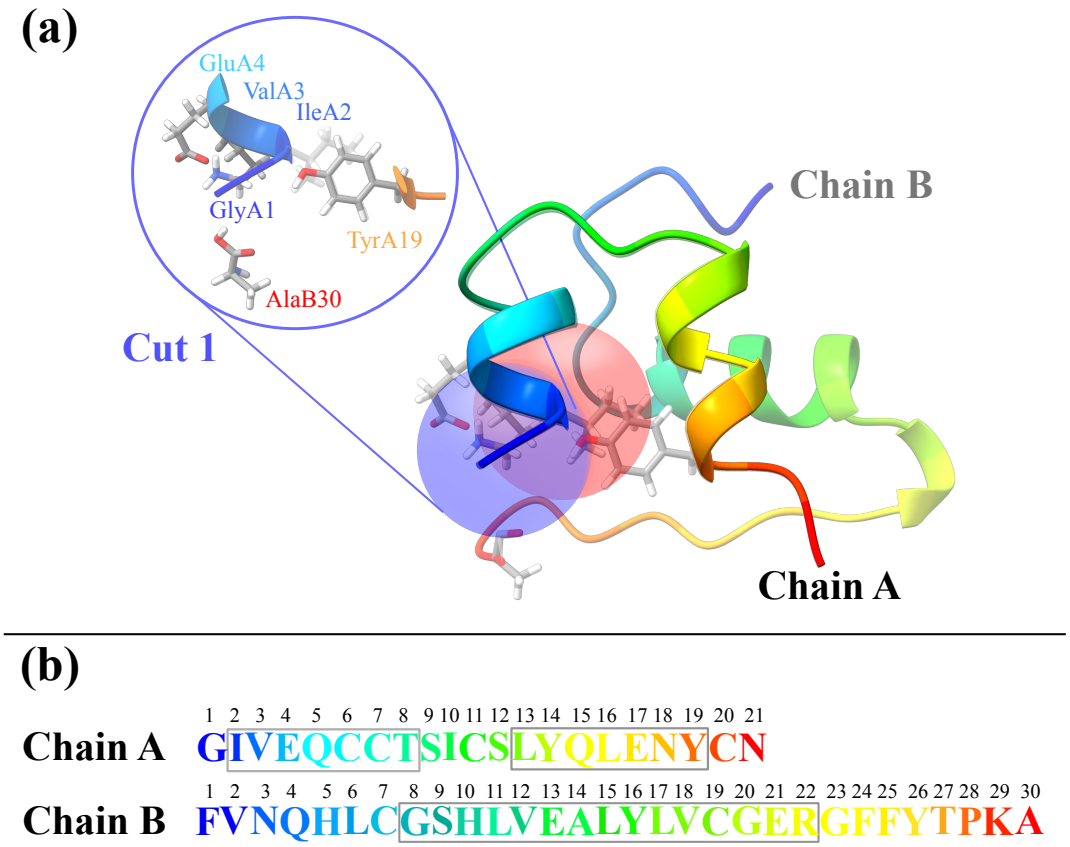
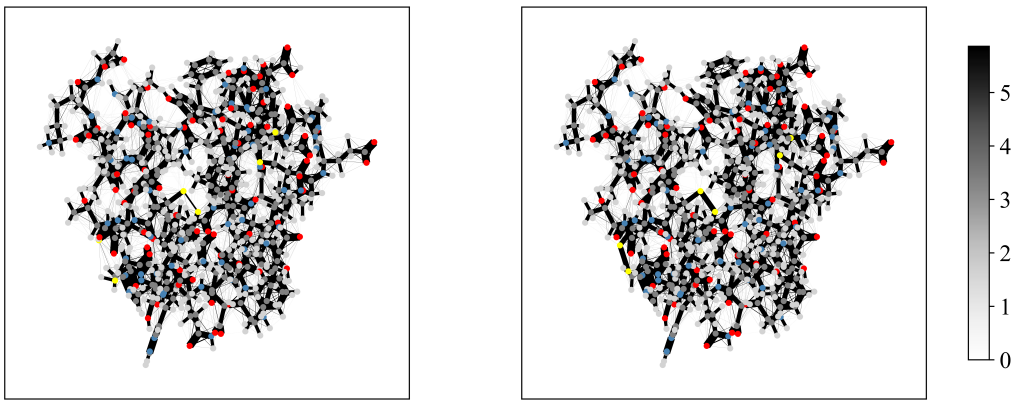


Figure 4.1: (a) Ribbon representation of the insulin protein structure, colored by residue index using a rainbow gradient. Two structural fragments (cuts) used in quantum calculations are highlighted as blue and red spheres, each with a 5.0 Å radius. The magnified inset depicts Cut 1, centered on the α-carbon of the first residue in Chain A (glycine), showing neighboring residues within the defined radius. Chain A is shown prominently in the foreground, while Chain B is rendered with fog in the background. The structure was visualized using ChimeraX [33]. (b) Primary amino acid sequences of Chain A (top, 21 residues) and Chain B (bottom, 30 residues), aligned with the color scheme used in panel (a).

To quantify electronic correlations in insulin, we first constructed a full MI matrix by stitching together the results from all overlapping spherical cuts. From this, we assembled the complete AMI matrix for the entire protein. This matrix was then visualized using a SPAWN (SPatial Atom-Wise Network) representation. Figure 4.2 compares the resulting cut-wise scheme 4.2a with a benchmark obtained from a direct full-protein DFT calculation 4.2b. In both networks, nodes represent atoms, and edges reflect non-zero AMI values. The edge thickness is scaled proportionally to the magni-

tude of AMI and measured in natural units (nats). As observed in our previous work [15], the network highlights strong electronic correlations along covalent bonds, with double bonds showing the greatest connectivity, followed by aromatic systems and then single bonds. Weaker non-covalent interactions are captured as thinner edges, highlighting the sensitivity of AMI to the underlying electronic structure. Both plots capture key electronic interactions, with the reference system exhibiting slightly more extensive non-covalent correlation. Nevertheless, the close correspondence between the two plots affirms the validity of the cut-wise reconstruction scheme.



(a) (b)

Figure 4.2: SPAWN plots of AMI across the insulin protein derived from (a) the cut-wise reconstruction of the full matrix and (b) a reference full-system DFT calculation. Nodes represent atoms and are color-coded by element: carbon (dark gray), hydrogen (light gray), oxygen (red), nitrogen (blue), sulfur (yellow). Edge thickness and gray gradient reflect pairwise AMI magnitude.

To extend the atomic-level insights to a more interpretable analysis of interactions between amino acid residues, we used a coarse-grained view of electronic correlations in insulin using FMI. Figure 4.3 compares SPAWN plots generated from the stitched AMI matrix obtained through fragment-wise DFT calculations with the one derived from the full-protein DFT calculation. These plots reveal strong correlations between cysteine pairs involved in disulfide bonds, Cys6–Cys11 (Chain A), Cys7A–Cys7B, and

Cys20A–Cys19B. While the cut-wise plot captures key local features, the reference plot exhibits additional long-range interactions absent in the stitched matrix due to fragment overlap limitations. Given that the same fragments will never co-occur in a single cut, this is to be expected. The comparison highlights the trade-off between locality and completeness in fragment-based correlation reconstructions.

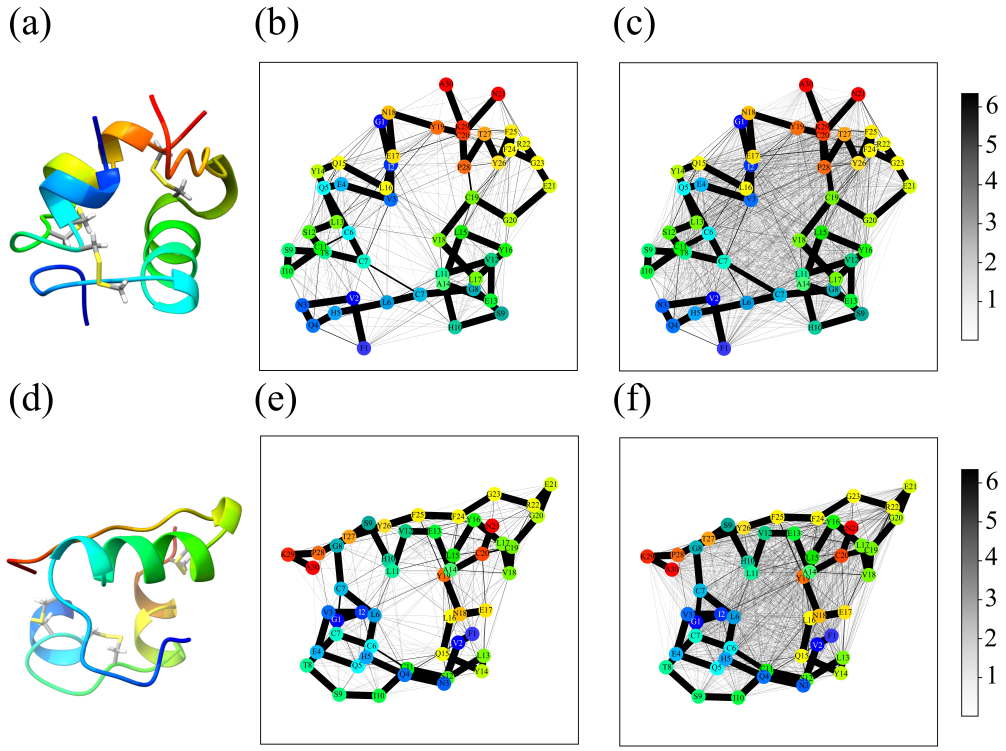


Figure 4.3: Coarse-grained visualization of electronic correlations in insulin using FMI. (a) Front and (d) side views of the insulin protein structure, with residues colored by sequence index using the rainbow gradient. Disulfide bonds are depicted as yellow sticks. (b) and (e) show SPAWN plots reconstructed from 49 overlapping structural fragments, while (c) and (f) present reference plots from a full-system DFT calculation. In all SPAWN plots, edges represent FMI values between residues, and edge thickness corresponds to correlation strength.

To evaluate the accuracy and completeness of the fragment-based FMI reconstruction, we compared heat maps derived from stitched and full-protein DFT calculations (Figure 4.4). Panel (a) shows the FMI matrix reconstructed from 49 overlapping fragments, while panel (b) displays the reference matrix computed from a full-system wavefunction. Chains A and B are enclosed in blue and green frames, respectively,

enabling intuitive separation of intra- and inter-chain interactions. Along the first off-diagonal, both matrices display strong FMI values corresponding to peptide bonds within the primary sequence.

More off-diagonal elements encode tertiary structure contacts. Among these, disulfide bridges between cysteine residues emerge as distinct violet patches, validating FMI's capacity to detect chemically significant covalent interactions.

The heat maps also resolve secondary structure segments. α-Helical regions are evident from periodic off-diagonal bands, reflecting canonical $i \to i+4$ hydrogen bonding of α-helices. These features, highlighted with purple frames, appear as regularly spaced darkening patterns along the diagonal. Their locations correspond closely with the hydrogen bonds listed in Table 4.1, which were identified using ChimeraX [33].

While classical tools like ChimeraX reliably detect standard hydrogen bonds, the FMI framework captures a broader range of electronic interactions—including subtle, non-classical, or transient correlations—that may not meet geometric criteria alone. As a result, the heat maps reveal additional interaction patterns not present in the table, underscoring the enhanced sensitivity and resolution of MI-based analysis.

These patterns align with the interactions reported in Table 4.1. The table lists 31 hydrogen bonds, with donor–acceptor distances between 2.7–3.3 Å, validating the signals observed in the heat map.

Despite the overall similarity between stitched and full DFT matrices, some interactions, such as the Glu17–Arg22 salt bridge, are absent from the stitched version due to their position outside overlapping fragments. This highlights a limitation of the fragment-based reconstruction strategy, which may be addressed by increasing the fragment radius or adapting alternative region selection. As the size of the cuts increases, similarity with the full DFT will improve at a computational cost.

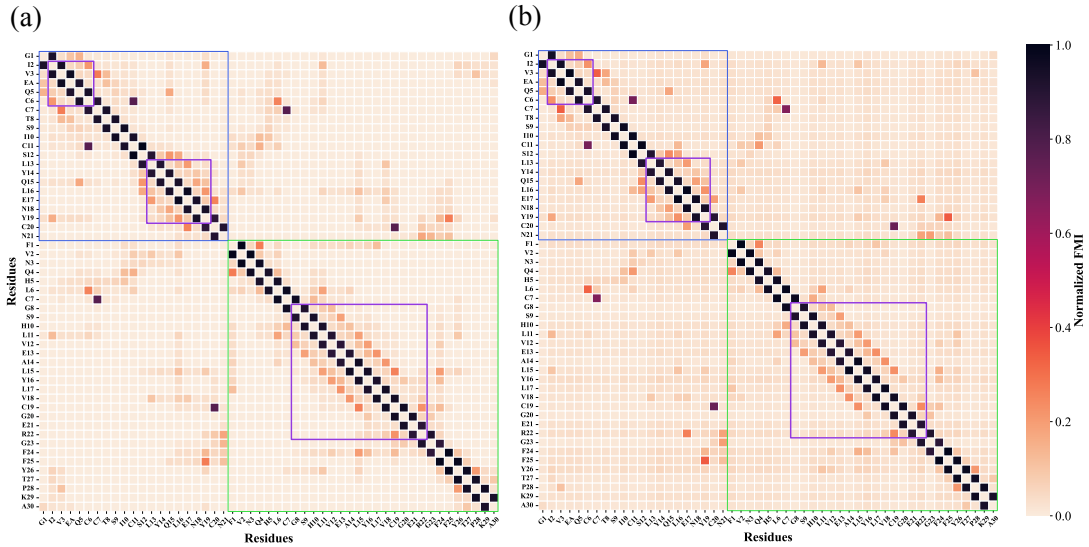


Figure 4.4: Heat maps of FMI across the insulin protein. (a) Reconstructed FMI matrix based on stitched DFT calculations from 49 overlapping structural fragments. (b) Reference FMI matrix computed from a full-protein DFT wavefunction. Blue and green frames delineate Chains A and B, respectively. Purple frames indicate α-helical regions exhibiting canonical $i \to i + 4$ hydrogen bonding.

No.	Donor	Acceptor	Dist. (Å)	No.	Donor	Acceptor	Dist. (Å)
1	GLN 5 (A)	GLY 1 (A)	3.168	17	LEU 6 (B)	CYS 6 (A)	2.723
2	CYS 6 (A)	ILE 2 (A)	2.828	18	HIS 10 (B)	CYS 7 (B)	3.150
3	CYS 7 (A)	VAL 3 (A)	2.824	19	LEU 11 (B)	GLY 8 (B)	3.132
4	THR 8 (A)	VAL 3 (A)	2.937	20	VAL 12 (B)	GLY 8 (B)	3.149
5	CYS 11 (A)	GLN 4 (B)	2.935	21	GLU 13 (B)	SER 9 (B)	2.887
6	GLN 15 (A)	SER 12 (A)	3.164	22	ALA 14 (B)	HIS 10 (B)	3.159
7	LEU 16 (A)	SER 12 (A)	2.998	23	LEU 15 (B)	LEU 11 (B)	2.977
8	GLU 17 (A)	LEU 13 (A)	3.003	24	TYR 16 (B)	VAL 12 (B)	2.923
9	ASN 18 (A)	GLN 15 (A)	3.215	25	LEU 17 (B)	GLU 13 (B)	2.974
10	TYR 19 (A)	LEU 16 (A)	3.001	26	VAL 18 (B)	ALA 14 (B)	2.897
11	CYS 20 (A)	GLU 17 (A)	3.234	27	CYS 19 (B)	LEU 15 (B)	2.915
12	ASN 21 (A)	GLY 23 (B)	2.968	28	GLY 20 (B)	TYR 16 (B)	3.222
13	GLN 4 (B)	PHE 1 (B)	2.889	29	ARG 22 (B)	CYS 19 (B)	3.256
14	HIS 5 (B)	CYS 7 (A)	3.112	30	ARG 22 (B)	CYS 19 (B)	3.319
15	ARG 22 (B)	GLU 17 (A)	3.183	31	GLY 23 (B)	GLY 20 (B)	3.122
16	PHE 25 (B)	TYR 19 (A)	2.998				

Table 4.1: List of 31 hydrogen bonds in insulin protein structure. Each entry includes donor and acceptor residues and their interatomic distances in (Å).

The AMI values obtained from the stitched matrix were compared against those from a full-protein DFT calculation (Figure 4.5). Each point in the scatter plot represents the AMI value between a pair of atoms. The data cluster closely along the identity line, indicating that the stitched matrix accurately reproduces electronic correlations derived from the full-system calculation. Gray points correspond to interactions absent from the stitched matrix; all of these have reference AMI values below 0.1 nat, suggesting that only electronically insignificant correlations are omitted. The overall distribution aligns with chemically meaningful AMI ranges: (i) non-bonded interac-

tions typically fall below 1 nat; (ii) single bonds between a heavy atom (C, O, N, or S) and hydrogen range from 2.0–2.5 nats; (iii) single bonds between heavy atoms span 2.7–4.7 nats; and (iv) double bonds exceed 4.8 nats. The presence of these expected clusters in both datasets confirms that the divide-and-correlate strategy not only preserves absolute correlation magnitudes but also maintains their underlying chemical hierarchy.

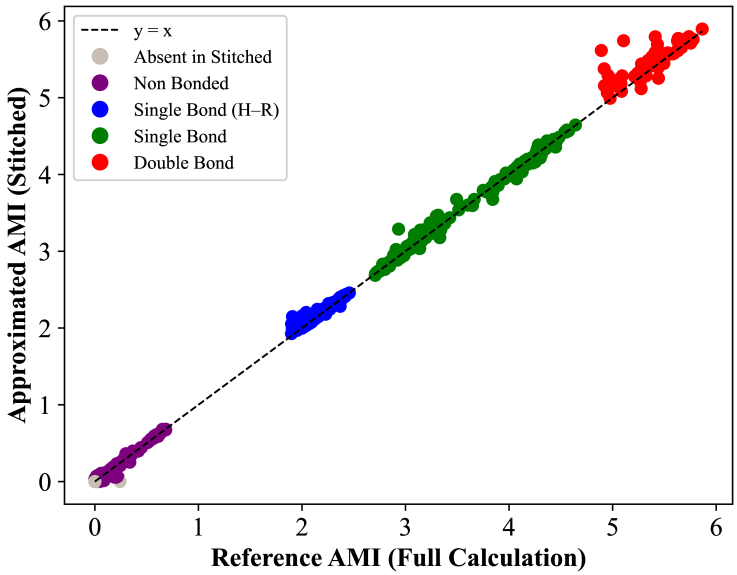


Figure 4.5: Scatter plot comparing AMI values from the stitched matrix with those from a full-protein DFT calculation for the insulin protein. Each point in the scatter plot represents the AMI value between a pair of atoms. Color-coded points correspond to distinct interaction types: gray points indicate interactions missing from the stitched matrix, purple for non-bonded interactions, blue for single bonds between a heavy atom and hydrogen, green for single bonds between heavy atoms, and red for double bonds. The black dashed line represents the identity line.

4.4 Conclusion

We have shown that AMI and FMI provide a meaningful framework for quantifying electronic correlations in proteins at multiple spatial resolutions. By adopting a fragment-based reconstruction strategy grounded in the principle of quantum locality, we mapped both interatomic and inter-residue correlations in insulin with accuracy

comparable to full-system DFT calculations, but at significantly reduced computational cost.

FMI captures a range of interactions, including covalent bonds, disulfide bridges, and hydrogen bonds, and highlights secondary structure elements like α-helices through distinct correlation patterns. Its agreement with classical hydrogen bonding data affirms its chemical validity, while its ability to uncover subtle, non-classical correlations underscores its broader utility.

The observed alignment between FMI and fragment overlap maps reinforces the nearsightedness of electronic interactions in proteins, validating the underlying assumption that spatial locality governs electronic correlation strength.

This work extends AMI beyond model peptides to realistic biomolecules, paving the way for quantum-informed exploration of protein–ligand interactions, pharmacophore modeling, and machine learning–driven drug discovery pipelines.

4.5 Acknowledgements

S.D.B acknowledges the NSERC Canada Research Chair and Discovery Grant program for financial support.

Bibliography

[1] A. Warshel, M. Levitt, Theoretical Studies of Enzymic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme, J. Mol. Biol. 103 (2) (1976) 227–249.

[2] A. J. Mulholland, Modelling Enzyme Reaction Mechanisms, Specificity and Catalysis, Drug Discov. Today 10 (20) (2005) 1393–1402.

[3] D. Baker, A. Sali, Protein Structure Prediction and Structural Genomics, Sci. 294 (5540) (2001) 93–96.

[4] H. M. Senn, W. Thiel, QM/MM Methods for Biomolecular Systems, Angew. Chem. Int. Ed. 48 (7) (2009) 1198–1229.

[5] M. Karplus, J. A. McCammon, Molecular Dynamics Simulations of Biomolecules, Nat. Struct. Biol. 9 (9) (2002) 646–652.

[6] U. Ryde, QM/MM Calculations on Proteins, Methods Enzymol. 577 (2016) 119–158.

[7] K. Raha, M. B. Peters, B. Wang, N. Yu, A. M. Wollacott, L. M. Westerhoff, K. M. Merz Jr, The Role of Quantum Mechanics in Structure-Based Drug Design, Drug Discov. Today 12 (17-18) (2007) 725–731.

[8] M. C. Ahmed, E. Papaleo, K. Lindorff-Larsen, How Well Do Force Fields Capture the Strength of Salt Bridges in Proteins?, PeerJ 6 (2018) e4967.

[9] F. Jensen, Introduction to Computational Chemistry, John wiley & sons, 2017.

[10] I. Shavitt, R. J. Bartlett, Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory, Cambridge university press, 2009.

[11] J. Rissler, R. M. Noack, S. R. White, Measuring Orbital Interaction Using Quantum Information Theory, Chem. Phys. 323 (2-3) (2006) 519–531.

[12] S. R. White, Density Matrix Formulation for Quantum Renormalization Groups, Phys. Rev. Lett. 69 (1992) 2863–2866.

[13] O. Legeza, R. M. Noack, J. Sólyom, L. Tincani, Applications of Quantum Information in the Density-Matrix Renormalization Group, in: Computational Many-Particle Physics, Vol. 739 of Lect. Notes Phys., Springer, Berlin/Heidelerg, 2008, pp. 653–664.

[14] G. K.-L. Chan, S. Sharma, The Density Matrix Renormalization Group in Quantum Chemistry, Annual review of physical chemistry 62 (1) (2011) 465–481.

[15] M. J. Moghadam, K. Boguslawski, R. Doucet, Ö. Legeza, P. Tecmer, S. De Baerdemacker, Correlation in Dipeptide Systems Form Quantum Information Theory Perspectives, chemrxiv. (2024).

[16] V. Timofeev, R. Chuprov-Netochin, V. Samigina, V. Bezuglov, K. Miroshnikov, I. Kuranova, X-Ray Investigation of Gene-Engineered Human Insulin Crystallized From a Solution Containing Polysialic Acid, Struct. Biol. Crystallogr. Commun. 66 (3) (2010) 259–263.

[17] R. A. DeFronzo, E. Ferrannini, Insulin Resistance: A Multifaceted Syndrome Responsible for NIDDM, Obesity, Hypertension, Dyslipidemia, and Atherosclerotic Cardiovascular Disease, Diabetes Care 14 (3) (1991) 173–194.

[18] T. Inaba, S. Tahara, N. Nisikawa, H. Kashiwagi, F. Sato, All-Electron Density Functional Calculation on Insulin With Quasi-Canonical Localized Orbitals, J. Comput. Chem. 26 (10) (2005) 987–993.

[19] B. Tah, P. Pal, S. Roy, D. Dutta, S. Mishra, M. Ghosh, G. Talapatra, Quantum-Mechanical DFT Calculation Supported Raman Spectroscopic Study of Some Amino Acids in Bovine Insulin, Spectrochim. Acta A Mol. Biomol. Spectrosc. 129 (2014) 345–351.

[20] T.-S. Lee, J. P. Lewis, W. Yang, Linear-Scaling Quantum Mechanical Calculations Of Biological Molecules: The Divide-And-Conquer Approach, Comput. Mat. Sci. 12 (3) (1998) 259–277.

[21] W. Yang, Electron Density As The Basic Variable: A Divide-And-Conquer Approach To The Ab Initio Computation Of Large Molecules, J. Mol. Struct.: THEOCHEM 255 (1992) 461–479.

[22] W. Yang, T.-S. Lee, A Density-Matrix Divide-And-Conquer Approach For Electronic Structure Calculations Of Large Molecules, J. Chem. Phys. 103 (13) (1995) 5674–5678.

[23] K. Boguslawski, P. Tecmer, Orbital Entanglement in Quantum Chemistry, Int. J. Quantum Chem. 115 (19) (2015) 1289–1295.

[24] K. Boguslawski, P. Tecmer, O. Legeza, M. Reiher, Entanglement Measures for Single-and Multireference Correlation Effects, J. Phys. Chem. Lett. 3 (21) (2012) 3129–3135.

[25] K. Boguslawski, P. Tecmer, G. Barcza, Ö. Legeza, M. Reiher, Orbital Entanglement in Bond-Formation Processes, J. Chem. Theory Comput. 9 (7) (2013) 2959–2973.

[26] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers, Softw. X 1 (2015) 19–25.

[27] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, GROMACS: Fast, Flexible, and Free, J. Comput. Chem. 26 (16) (2005) 1701–1718.

[28] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, J. Am. Chem. Soc. 118 (45) (1996) 11225–11236.

[29] W. L. Jorgensen, J. Tirado-Rives, The OPLS [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin, J. Am. Chem. Soc. 110 (6) (1988) 1657–1666.

[30] F. Neese, The ORCA Program System, WIREs Comput. Mol. Sci. 2 (1) (2012) 73–78.

[31] N. Mardirossian, M. Head-Gordon, ΩB97M-V: A Combinatorially Optimized, Range-Separated Hybrid, Meta-Gga Density Functional With VV10 Nonlocal Correlation, J. Phys. Chem. 144 (21) (2016) 214110.

[32] R. Ditchfield, W. J. Hehre, J. A. Pople, Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules, J. Chem. Phys. 54 (2) (1971) 724–728.

[33] E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Tools for Structure Building and Analysis, Protein Sci. 32 (11) (2023) e4792.

Chapter 5

Discussion

In recent years, quantum information theory has gained increasing traction in quantum chemistry as a powerful framework for quantifying electronic correlations. A particularly insightful avenue has been the use of MI derived from reduced density matrices, which allows the assessment of entanglement between pairs of orbitals. This thesis introduces a novel application of this concept by extending MI to the atomic and fragment levels, namely AMI and FMI, thereby providing spatially localized insight into electronic correlations in biomolecular systems.

5.1 From Orbitals to Atoms

Pioneering studies by Rissler et al. [1] and Legeza et al. [2] utilized orbital-based MI, derived from two-orbital reduced density matrices, to analyze correlation patterns in strongly correlated systems and conjugated molecules. These works established the utility of MI for mapping electronic interactions but relied primarily on canonical or localized MOs, which often obscure spatial locality. Szalay and co-workers [3] advanced this framework by introducing entanglement measures for chemical bonding, providing deeper insight into static and dynamic correlations.

Canonical molecular orbitals, being inherently delocalized, obscure chemically intu-

itive patterns of electron correlation in terms of local entities. Localized orbitals like Pipek-Mezey [4] or Foster-Boys [5] improve interpretability but still retain some molecular character, making it difficult to assign them to individual spatial centers. Our approach distinguishes itself by operating directly on the AO basis, providing a spatially confined and more transparent framework. AMI patterns in this way align closely with atomic identities and preserve locality across diverse systems. These findings are consistent with the recent analysis by Ding et al. [6], noting that entanglement should be distributed across separated local regions and accessed via orbitals localized to as few atomic centers as possible.

While prior studies [7, 8] utilized circular diagrams to depict MI, our SPAWN plots extend this concept to an atom-centered framework that preserves the real-space geometry of the molecule. The SPAWN visualizations introduced in this study serve as a powerful tool for analyzing the spatial structure of electronic correlations. By encoding AMI values into the thickness of edges connecting atomic centers, these spatially resolved plots highlight both primary bonding regions and subtle interactions.

To validate our summation-based AMI approximation, we explicitly compared it against full block-wise MI values obtained from high-accuracy DMRG calculations on small molecular systems such as hydrogen cyanide (HCN) and formaldehyde (CH_2O). The strong agreement observed in these comparisons demonstrates that our approximations capture the essential interatomic correlation patterns while avoiding the computational cost of full subsystem tracing. The main difference to block-wise MI, which requires repeated orbital reordering and expensive tensor contractions, is that our summation-based AMI is computationally lighter, easier to interpret, and scalable to larger biomolecular systems. This design reflects the broader motivation expressed by Aliverti-Piuri et al. [9], emphasizing the importance of scalable and operationally meaningful correlation measures for integrating QIT into practical quantum chemical applications.

By extending the analysis to 400 dipeptides, we evaluated how the side-chains of amino acids influence electronic correlation. The goal of this exhaustive sampling was to construct a representative library of all possible pairwise interactions between neighboring amino acids, which can later serve as training data for machine learning models aimed at predicting entanglement patterns in larger proteins. The results revealed that covalent bonds exhibited the highest AMI, followed by weaker hydrogen-involving interactions, with non-bonded interactions showing the lowest values. This result reinforces the foundational principle that valence electrons dominate entanglement in molecular systems, a concept emphasized by Ding and coworkers [10] where they demonstrated that entanglement patterns derived from Maximally Entangled Atomic Orbitals (MEAOs) effectively recover both traditional Lewis (two-center) and more complex multicenter bonding structures. This indicates that the entanglement captured by MEAOs corresponds closely with chemically active regions, which are predominantly associated with valence electrons.

Bond stretching in the Gly-Gly dipeptide was employed to probe the sensitivity of AMI. As the peptide bond was progressively elongated, a clear decrease in AMI was observed, reflecting the weakening of electronic correlation across the bond. This decay trend is consistent with findings in the literature. For instance, Ding et al. [11] introduced a rigorous framework for orbital entanglement, identifying similar decay behaviors between entangled orbitals based on the spatial and symmetry properties of molecular systems. Likewise, Zhao et al. [7] employed DMRG-based MI to monitor changes in electronic structure during bond formation and cleavage in metal–ethene complexes. Their study demonstrated that MI serves as a reliable descriptor for the evolving correlation between metal centers and ligands. These studies reinforce the broader applicability of MI in tracking bond reorganization and correlation decay across a range of chemical systems.

One limitation of our approach stems from its reliance on the single Slater deter-

minant wavefunction. This simplification was necessitated by the computational challenges associated with constructing higher-order RDMs required for multireference treatments. In systems described by multireference wavefunctions, the application of Wick's theorem 1.41 is no longer valid, complicating the direct evaluation of MI. While our benchmark comparisons with DMRG demonstrate that HF and DFT methods can reproduce the qualitative features of AMI with reasonable fidelity, the single-determinant approaches inherently fail to capture dynamic excitation effects and static (nondynamic) correlation. As a result, in strongly correlated systems or those exhibiting near-degeneracies, this limitation may lead to a systematic underestimation of interatomic correlations and entanglement strength.

Computational comparisons across RHF, DFT (B3LYP and ωB97M-V), and DMRG demonstrated that while MI magnitudes vary due to methodological differences, the correlation topologies remain qualitatively similar. This supports the findings of Stein and Reiher [12], who noted that even with modest computational settings entanglement remains sufficiently accurate, even when the total energy is not fully converged. Therefore, even a single-reference wavefunctions can still yield meaningful MI maps that reflect the dominant orbital interactions. These MI patterns can be effectively used to determine the optimal ordering of orbitals in MPS representations for DMRG calculations, thereby accelerating convergence and enhancing numerical stability. Legeza and Sólyom [13] demonstrated that the ordering of orbitals according to their entropic content significantly influences the convergence and interpretability of DMRG. By placing orbitals with high von Neumann entropy near the center of the chain, quantum information exchange between blocks is maximized, leading to more entangled and efficiently represented wavefunctions.

Building on this idea, Barcza et al. [14] introduced a graph-theoretical orbital ordering scheme using the Fiedler vector, derived from the Laplacian of an MI graph. In this graph, each node corresponds to an orbital, and edges are weighted by the two-orbital

MI, I_ij. The Fiedler vector, which is the eigenvector corresponding to the second smallest eigenvalue of the graph Laplacian, provides a one-dimensional embedding of the orbital network. By sorting orbitals according to their Fiedler vector components, strongly correlated orbitals are placed adjacently in the DMRG chain, minimizing the entanglement distance and enhancing the efficiency of the wavefunction representation. This approach has since been widely adopted for entanglement-aware orbital ordering in tensor network methods.

Another important consideration is the basis set dependence of AMI calculations. This dependence was evaluated in the Cys-Gly system through a basis set convergence study using the cc-pVXZ series. Although the AMI initially increases with basis sets size, it began to plateau gradually. This behavior indicates that AMI converges if the basis set becomes complete. According to the framework proposed by Aliverti-Piuri et al. [9], QIT differentiates between particle correlation which is basis independent and orbital correlation, which is basis dependent. The observed growth in AMI with increasing basis set size is therefore expected, as it corresponds to the refinement of orbital resolution within an expanding Hilbert space. While methods like QTAIM [15] offers a basis-set independent framework grounded in electron density topology, its conclusions can depend on arbitrary threshold values and may diverge from conventional bonding concepts. By comparison, AMI is rooted in quantum information theory, offering a systematic, mathematically grounded, and chemically intuitive approach to understanding electronic interactions.

The importance of entropic measures in chemistry is increasingly recognized. As emphasized by Aliverti-Piuri et al. [9], entanglement at the orbital level provides a nuanced view of correlation complexity, complementing traditional energy-based descriptors and enabling a more complete characterization of electronic structure. Our work builds on this by extending entanglement diagnostics to the atomic level, integrating chemical granularity into quantum correlation metrics.

A key advantage of the AMI framework lies in its foundation on RDMs, which enables a rigorous, unbiased analysis of electronic structure. Unlike EDA [16, 16], which requires user-defined molecular fragments, AMI quantifies the total (quantum and classical) correlation directly from the wavefunction. However, AMI quantifies information-based correlations without direct energetic interpretation. While we observed that AMI trends often correlate well with EDA values, AMI does not substitute for energetic analyses. This distinction may limit AMI interpretability in applications where energetic driving forces are of primary interest.

AMI does not rely on assumptions about what constitutes a bond or on predefined donor–acceptor roles. In contrast, NBO [17] analysis depends on orbital localization schemes that are inherently non-unique and sensitive to the chosen method, potentially leading to varying interpretations.

5.2 From Atoms to Residues

By transitioning from atomic-level MI to a coarse-grained level we introduced the FMI 3.8. This approach enables the study of electronic correlations not just between atoms, but between chemically meaningful subunits like amino acid residues. By aggregating AMI contributions across amino acids subunits, we constructed a residue-level representation of electronic interaction. This abstraction is particularly well-suited for biomolecules, where structure and function are often organized at the level of residues or domains rather than individual atoms.

The FMI was compared with EDA across various dipeptides. While the EDA in a study by Levine and Head-Gordon [18] offers a detailed partitioning of chemical bond energies into components like spin-coupling and charge transfer, our approach using AMI/FMI provides a complementary perspective by quantifying the total electronic correlation between atoms or fragments, encompassing both classical and quantum

contributions. This allows us to capture nuances of electronic interactions, such as coherence and entanglement, that are not directly addressed by traditional EDA methods.

We observed conformations with similar energies can have different FMI values. This underscores the fact that FMI is a measure of correlation strength, not necessarily energetic favorability. Elevated FMI between two residues may reflect transient or constrained geometries that are electronically entangled, even if energetically unfavorable. Thus, FMI should not be viewed as a surrogate for EDA or the Lennard-Jones component of the force fields. Instead, it offers a distinct layer of information which can guide interpretations of interaction networks and folding pathways with a different lens.

SPAWN visualizations of FMI revealed spatially resolved patterns of inter-residue correlation. These maps provided immediate visual insight into which regions of a peptide or protein are most electronically correlated, complementing classical representations based on hydrogen bonding or proximity. Unlike traditional metrics that rely solely on geometry or electrostatics, FMI integrates quantum entanglement into the interpretation of molecular structure. This gives rise to a chemically rich and topologically grounded view of interaction networks.

The concept of coarse-grained MI parallels the block entropy approach of Legeza and Sólyom [13], which quantifies subsystem entanglement in DMRG. However, our FMI method is significantly more scalable and avoids the costly subsystem optimization and tensor contractions required by block entropy, enabling efficient application to large biomolecular systems where DMRG is otherwise impractical. FMI reveals inter-residue correlation patterns that highlight electronically coupled regions across biomolecules. These patterns go beyond classical metrics such as hydrogen bonding, partial charge interactions, or hydrophobic contacts, offering a fundamentally different layer of insight grounded in wavefunction correlations. The use of dynamic FMI

in MD simulations of the Neh2 peptide (Chapter 3) underscored how electronic interactions evolve during folding. FMI captures residue-residue entanglement patterns that correlate with stabilizing interactions. It not only identifies chemically relevant interactions—such as hydrogen bonds, salt bridges, and even non-polar contacts but also quantifies their relative strength via the magnitude of MI. This level of resolution enables a more nuanced evaluation of stabilizing interactions, distinguishing, for example, between weak and transient hydrogen bonds versus more persistent, stronger ones. Such differentiation is essential in contexts like protein folding, allosteric regulation, and drug binding, where the strength and persistence of specific contacts often dictate biological function.

One promising avenue for FMI lies in its integration with ML. Existing ML-based methods such as ANI, AIMNet, and PhysNet have demonstrated impressive accuracy by learning from quantum chemical data. ANI (Accurate NeurAl networK engINe) uses atom-centered symmetry functions to encode local geometric environments and predict molecular energies with near-DFT accuracy at low cost [19]. AIMNet (Atoms-in-Molecules Neural Network) extends this by incorporating both local and global context using message-passing and attention-like mechanisms, enabling multitask prediction of energies, charges, and spins [20]. PhysNet is another physics-inspired architecture that predicts not only energies and forces but also dipole moments and partial charges, incorporating physical constraints such as energy conservation and rotational equivariance [21]. However, all these models primarily rely on geometric or energy-based descriptors and do not explicitly account for quantum correlations or entanglement. Incorporating FMI as a feature would introduce an entropic, correlation-based descriptor that captures nonlocal quantum mechanical interactions, offering complementary information and potentially enhancing predictive accuracy, especially for complex biomolecular systems.

5.3 Toward Larger Scales

A central objective of this study was to make quantum correlation analysis feasible for large biomolecular systems. While MI has been effectively employed in small molecules to reveal electronic correlations and bonding patterns, its application to full proteins remains computationally intractable due to the computational cost for the full-system wavefunction.

To address this limitation, we developed a cut-wise reconstruction strategy that circumvents the need for a full-system wavefunction. By calculating RDMs on overlapping spatial fragments and stitching their contributions together, we were able to approximate the full-system FMI matrix with high accuracy. This approach not only offers linear computational scaling with system size but also maintains key features of the global correlation structure. This approach, derived from DFT is, in principle, still exponential in cost, but becomes practically tractable when confined to small overlapping fragments. This enables the method to be applied to large biomolecules with hundreds of atoms, vastly extending the reach of quantum correlation analysis in structural biology and computational chemistry. It opens possibilities for studying systems that are far beyond the capabilities of DMRG or other wavefunction-based techniques.

The success of our approach hinges on the locality of electronic correlation. This nearsightedness behavior justifies the use of spatially localized fragments to approximate global correlation structure without incurring substantial information loss.

We demonstrated the feasibility and robustness of this methodology using the insulin protein in chapter 4, which contains a variety of structural motifs including α-helices, disulfide bridges, and inter-residue hydrogen bonds. The stitched FMI matrix reconstructed from spatial fragments closely mirrored the full-system FMI calculated directly from a global DFT wavefunction. Importantly, chemically meaningful features such as characteristic α-helical regions with $i \rightarrow i+3$ hydrogen bond patterns

emerged clearly in the reconstructed map.

In a few cases, some interactions were missed in the reconstructed FMI matrix. The primary reason for these missing terms is that they are not sampled in any single spatial cut. In this context, fragmentation boundaries remain the primary source of information loss. Nevertheless, increasing the fragment radius beyond 5 Å, or dynamically adapting fragment boundaries based on interaction density, could reduce information loss. However, doing so also increases the computational cost per fragment, resulting in a trade-off between accuracy and efficiency. This challenge is analogous to known limitations in fragmentation-based energy methods, where interactions between distant fragments must be carefully reconstructed to maintain accuracy [22, 23, 24, 25].

In conclusion, FMI expands the toolkit of quantum chemical descriptors, providing an interpretable, scalable, and classical and quantum correlation measure. It bridges the orbital and atomistic representations developed in chapter 2 with biologically relevant scales, enabling the application of QTI to large biomolecular systems.

5.4 Future works

A promising extension of this framework would involve replacing orthogonalized AOs with domain-localized orbitals (DLOs) such as those employed in DLPNO methodologies [26, 27, 28]. Incorporating AMI into the DLO framework could significantly enhance the spatial resolution and chemical specificity of correlation maps. It would enable a more refined dissection of electronic interactions within catalytically active sites, enzyme pockets, or protein–ligand interfaces—regions where traditional orbital partitions often fail to provide adequate granularity. Moreover, using DLOs could help mitigate some of the basis set dependence observed with AO-based AMI, as the localization criteria used in DLO generation are more robust across basis set hierar-

chies. This could improve the transferability and comparability of AMI results across different chemical systems and computational protocols.

This technique complements QM/MM schemes by offering a pre-screening tool for identifying regions requiring quantum treatment. It could inform adaptive QM/MM partitions based on correlation density rather than classical descriptors. Beyond peptides, FMI could be extended to nucleic acids, capturing base pairing and stacking interactions. The method could also map protein-ligand or protein–nucleic acid interactions, offering a QIT-based fingerprint of binding interfaces. For example, $\pi - \pi$ stacking between DNA bases or aromatic drug residues could be detected and quantified through FMI.

In addition, integrating FMI with structural or sequence-based descriptors can further enhance its utility in machine learning applications. Unlike classical features such as bond lengths or angles, the FMI matrix also captures inherently quantum mechanical aspects of inter-residue interactions. This matrix provides unique information about electronic correlations that can be directly incorporated into ML workflows. For instance, in graph neural networks (GNNs), residues or atoms can be represented as nodes, with edge weights informed by FMI values. This allows the model to encode correlation-driven chemical environments, potentially improving performance in tasks such as protein structure prediction, ligand binding affinity estimation, and residue interaction mapping.

A further extension would be the integration of dynamic FMI analysis with *ab initio* MD. This approach could provide a uniquely quantum mechanical perspective on how electronic correlations evolve over time during processes such as conformational transitions, or ligand binding. By capturing the temporal dynamics of entanglement, this method would offer deeper insight into the electronic underpinnings of biologically relevant structural changes.

Another promising direction lies in training a ML model to predict FMI matrices

from fragment-level data. By compiling a sufficiently large dataset of proteins with full and stitched FMI matrices across diverse fragment sizes and structures, one could train a model to obtain a predicted FMI matrix with almost no cost.

It is also worth investigating the sensitivity of FMI to the choice of density functional and basis set. While ωB97M-V performed well in the present work, systematic benchmarking across different methods and basis sets should inform best practices for FMI-based analysis. This is particularly important if the approach is to be applied to a wide range of protein classes or unusual chemical environments.

Bibliography

[1] J. Rissler, R. M. Noack, S. R. White, Measuring Orbital Interaction Using Quantum Information Theory, Chem. Phys. 323 (2006) 519–531.

[2] Ö. Legeza, R. M. Noack, J. Sólyom, L. Tincani, Applications of Quantum Information in the Density-Matrix Renormalization Group, in: Computational Many-Particle Physics, Vol. 739 of Lect. Notes Phys., Springer, Berlin/Heidelerg, 2008, pp. 653–664.

[3] S. Szalay, G. Barcza, T. Szilvási, L. Veis, Ö. Legeza, The Correlation Theory of the Chemical Bond, Sci. Rep. 7 (1) (2017) 2237.

[4] J. Pipek, P. G. Mezey, A Fast Intrinsic Localization Procedure Applicable for Ab-initio and Semiempirical Linear Combination of Atomic Orbital Wave Functions, J. Chem. Phys. 90 (1989) 4916–4926.

[5] J. Foster, S. Boys, Canonical Configurational Interaction Procedure, Rev. Mod. Phys. 32 (1960) 300.

[6] L. Ding, G. Dünnweber, C. Schilling, Physical Entanglement Between Localized Orbitals, Quantum Sci. Technol. 9 (1) (2023) 015005.

[7] Y. Zhao, K. Boguslawski, P. Tecmer, C. Duperrouzel, G. Barcza, Ö. Legeza, P. W. Ayers, Dissecting the Bond-Formation Process of D 10-Metal–ethene Complexes With Multireference Approaches, Theor. Chem. Acc. 134 (10) (2015) 120.

[8] S. Szalay, M. Pfeffer, V. Murg, G. Barcza, F. Verstraete, R. Schneider, Ö. Legeza, Tensor Product Methods and Entanglement Optimization for Ab Initio Quantum Chemistry, Int. J. Quantum Chem. 115 (2015) 1342–1391.

[9] D. Aliverti-Piuri, K. Chatterjee, L. Ding, K. Liao, J. Liebert, C. Schilling, What Can Quantum Information Theory Offer to Quantum Chemistry?, Faraday Discuss. 254 (2024) 76–106.

[10] L. Ding, E. Matito, C. Schilling, From Entanglement to Bonds: Chemical Bonding Concepts From Quantum Information Theory, arXiv preprint arXiv:2501.15699 (2025).

[11] L. Ding, C. Schilling, Correlation Paradox of the Dissociation Limit: A Quantum Information Perspective, J. Chem. Theory Comput. 16 (7) (2020) 4159–4175.

[12] C. J. Stein, M. Reiher, Automated Selection of Active Orbital Spaces, J. Chem. Theory Comput. 12 (4) (2016) 1760–1771.

[13] O. Legeza, J. Sólyom, Optimizing the Density-Matrix Renormalization Group Method Using Quantum Information Entropy, Phys. Rev. B 68 (2003) 195116.

[14] G. Barcza, Ö. Legeza, K. H. Marti, M. Reiher, Quantum-Information Analysis of Electronic States of Different Molecular Structures, Phys. Rev. A 83 (1) (2011) 012508.

[15] R. F. W. Bader, Atoms in Molecules: A Quantum Theory, Oxford University Press, 1990.

[16] M. v. Hopffgarten, G. Frenking, Energy Decomposition Analysis, WIREs Comput. Mol. Sci. 2 (1) (2012) 43–62.

[17] E. D. Glendening, C. R. Landis, F. Weinhold, Natural Bond Orbital Methods, WIREs Comput. Mol. Sci. 2 (1) (2012) 1–42.

[18] D. S. Levine, M. Head-Gordon, Energy Decomposition Analysis of Single Bonds Within Kohn–Sham Density Functional Theory, Proc. Natl. Acad. Sci. 114 (48) (2017) 12649–12656.

[19] J. S. Smith, O. Isayev, A. E. Roitberg, ANI-1: An Extensible Neural Network Potential With DFT Accuracy at Force Field Computational Cost, Chem. Sci. 8 (2017) 3192–3203.

[20] R. Zubatyuk, J. S. Smith, J. Leszczynski, O. Isayev, Accurate and Transferable Multitask Prediction of Chemical Properties With an Atoms-in-Molecules Neural Network, Science Advances 5 (8) (2019) eaav6490.

[21] O. T. Unke, M. Meuwly, PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges, Journal of Chemical Theory and Computation 15 (6) (2019) 3678–3693.

[22] M. A. Addicoat, M. A. Collins, Accurate Treatment Of Nonbonded Interactions Within Systematic Molecular Fragmentation, J. Chem. Phys. 131 (10) (2009).

[23] N. Jiang, J. Ma, Y. Jiang, Electrostatic Field-Adapted Molecular Fractionation With Conjugated Caps For Energy Calculations Of Charged Biomolecules, J. Chem. Phys. 124 (11) (2006).

[24] D. W. Zhang, J. Zhang, Molecular Fractionation With Conjugate Caps For Full Quantum Mechanical Calculation Of Protein–Molecule Interaction Energy, J. Chem. Phys. 119 (7) (2003) 3599–3605.

[25] X. Chen, D. Zhang, J. Zhang, Fractionation Of Peptide With Disulfide Bond For Quantum Mechanical Calculation Of Interaction Energy With Molecules, The Journal of chemical physics 120 (2) (2004) 839–844.

[26] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese, Sparse Maps—A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory, J. Chem. Phys. 144 (2) (2016) 024109.

[27] E. Francisco, A. M. Pendás, Towards an Energy Partition Into Real Space Resonance Structures: 1-and 2-Particle Density Matrix Decomposition, Mol. Phys. 114 (7-8) (2016) 1334–1344.

[28] E. Francisco, A. M. Pendás, Electron Number Distribution Functions From Molecular Wavefunctions. Version 2, Comput. Phys. Commun. 185 (10) (2014) 2663–2682.

Vita

Mostafa Javaheri Moghadam

Education

PhD | University of New Brunswick | Fredericton, NB, Canada — 2026
Major: Chemistry

Master of Science | Golestan University | Iran (Recognized as equivalent by World Education Services) — 2015
Major: Biochemistry

Bachelor of Science | University of Environment | Iran (Recognized as equivalent by World Education Services) — 2011
Major: Natural Resources Engineering-Environment

Certifications

- Business Immigrant Essentials Program, Ignite Fredericton (Nov 2025).
- Invention to Innovation (i2I) Program, Beedie School of Business, Simon Fraser University (Sep 2025)
- BioInnovation Scientist (BIS)-Foundation Program, AdMare BioInnovations (Nov 2024).

- Application of Molecular Dynamics Simulation in Structural Biology, Institute of Biochemistry and Biophysics (IBB) (June 2022).
- Introduction to Flow Cytometry, Institute of Biochemistry and Biophysics (IBB) (May 2022).
- Quantitative Structure Activity Relationship (QSAR), Institute of Biochemistry and Biophysics (IBB) (May 2014).

Awards

- Lab2Market Launch Program & Mitacs Business Strategy Internship (BSI), Oct 2025: Funded commercialization and market-entry strategy for LigandQI.
- Lab2Market Validate program & Mitacs BSI, Aug 2024: Awarded for LigandQI, a software focused on optimizing drug design by analyzing protein-ligand interactions.
- Recipient of the 2023 Karel Wiesner Postgraduate Scholarship in Chemistry for demonstrating both academic and research excellence chemistry
- Mitacs Globalink Research Award (GRA) 2023.
- Distinguished presentation in Young Science Beyond Borders (YSBB) 2023.
- Recipient of the 2022 Frank J. and Norah Toole Graduate Scholarship in Chemistry for demonstrating both academic and research excellence.

Publications

- **Javaheri Moghadam, M.**, Mulder, R., De Baerdemacker, S., Divide and Correlate: Mapping Electronic Correlations in Proteins via Local Cut-Wise Reconstruction. Chem. Commun. 2026.

- **Javaheri Moghadam, M**, Boguslawski, K, Tecmer, P, De Baerdemacker, S, 2025. "From dipeptide systems to polypeptides: evolution of mutual information." *ChemRxiv*

- **Javaheri Moghadam, M**, Boguslawski, K, Doucet, R, Legeza, O, Tecmer, P, De Baerdemacker, S, 2024, Correlation in Dipeptide Systems form Quantum Information Theory Perspectives, ChemRxiv, DOI: 10.26434/chemrxiv-2024-2qs86.

- Noroozi Tiyoula, F, Aryapour, H, **Javaheri Moghadam, M**, 2022, Comparative study of the unbinding process of some HTLV-1 protease inhibitors using unbiased molecular dynamics simulations, PLOS ONE 17(7), DOI: 10.1101/2022.01.17.476663.

- Sohraby, F, **Javaheri Moghadam, M**, Aliyar, M, & Aryapour, H, 2021, Complete reconstruction of Dasatinib unbinding pathway from the c-Src kinase by supervised molecular dynamics simulation method; assessing efficiency and trustworthiness of the method, Journal of biomolecular structure & dynamics, DOI: 10.1080/07391102.2021.1972839, 1 – 11.

- Sohraby, F, **Javaheri Moghadam, M**, Aliyar, M, & Aryapour, H, 2020, A boosted unbiased molecular dynamics method for predicting ligands binding mechanisms: Probing the binding pathway of Dasatinib to the Src-kinase, Bioinformatics, DOI: 10.1093/bioinformatics/btaa565, 1367-4803.

- Sohraby, F, Hadi Soltanabad, M, Bagheri, M, Bezi Javan, M, **Javaheri Moghadam, M**, Kohan Baghkheirati, E, Bagherieh Najjar, M, 2020, Application of molecular dynamics in coating Ag-conjugated nanoparticles with potential therapeutic applications, Nano Biomedicine and Engineering 12 (1), 90-98.

- Sohraby, F, Bagheri, M, **Javaheri Moghadam, M** & Aryapour, H, 2017, In silico prediction of new inhibitors for the nucleotide pool sanitizing enzyme, MTH1, using drug repurposing, Journal of biomolecular structure & dynamics, 1-9.

Presentation

- Javaheri Moghadam, M, Mulder, R, De Baerdemacker, S. Quantifying Electronic Correlations in Insulin via Mutual Information: A Cut-Wise, Quantum-Local Approach. ResearchNB Health Symposium 2025.

- Javaheri Moghadam, M. De Baerdemacker, S. Quantum Mutual Information in Biomolecular Systems, Canadian Chemistry Conference and Exhibition (CSC Ottawa), 2025.

- Javaheri Moghadam, M. De Baerdemacker, S. Transforming Drug Discovery Through Quantum Intelligence, New Brunswick Health Research Symposium, 2024.

- Javaheri Moghadam, M. De Baerdemacker, S. Quantum Correlation in Dipeptide Systems: A Quantum Information Theory Approach, Canadian Symposium On Theoretical And Computational Chemistry, 2024.

- Javaheri Moghadam, M. De Baerdemacker, S. Quantum mutual information: a novel insight into amino acid interactions and chemical bonding in biomolecules, Conference of the Biological Sciences, 2024.

- Javaheri Moghadam, M. De Baerdemacker, S. Deciphering Amino Acid Interactions: Insights from Quantum Mutual Information Analysis in Biomolecular Systems, GRC Research Brief Challenge, 2024.

- Javaheri Moghadam, M. De Baerdemacker, S. Quantum Mutual Information: A Novel Insight into Amino Acid Interactions and Chemical Bonding in Biomolecules, Maritimes Biological Chemistry Conference (MBCC), 2023.

- Javaheri Moghadam, M. De Baerdemacker, S. Measuring local correlations in biomolecular systems using quantum information theory, 17th International Congress of Quantum Chemistry (ICQC), Bratislava, 2023 https://icqc2023.org/ .

- Javaheri Moghadam, M. De Baerdemacker, S. Exploring Molecular Bonding through Mutual Information: A Quantum Information Approach, Young Science Beyond Borders (YSBB) 2023, https://amu.pan.pl/ysbb .

- Javaheri Moghadam, M. De Baerdemacker, S. Measuring local correlation in extended biomolecular systems from information theory measures, 28th UNB Graduate Research Conference 2023.

- Javaheri Moghadam, M. De Baerdemacker, S. Local entanglement measurement in macro molecules, 12th Triennial Congress of the World Association of Theoretical and Computational Chemists, www.watoc2020.ca .

- Javaheri Moghadam, M. De Baerdemacker, S. Measuring local correlation in extended molecular systems, 27th UNB Graduate Research Conference 2022.

- Javaheri Moghadam, M. De Baerdemacker, S. Objective Incorporating quantum correlation in extended molecular systems using advanced electronic structure methods in quantum chemistry, IUPAC CCCE 2021 - 48th World Chemistry Congress & 104th Canadian Chemistry Conference and Exhibition.

- Javaheri Moghadam, M, De Baerdemacker, S. Incorporating quantum correlation in extended molecular systems using advanced electronic struc-

ture methods in quantum chemistry, 2021 #RSCPoster Twitter Conference, https://twitter.com/MjJavaheri/status/1366722523329687559.

- Javaheri Moghadam, M, Aryapour, H, Dehno Khalaji, A, Designing new 4h-chromene compounds with anticancer activity and studying their interaction with tubulin using molecular docking approach, Iranian. J .Immunol. 11 (Supplement1): p. 151-152, Poster presentation.