Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Matematica

Molecular docking via quantum annealing: modelling $\pi-$stacking interactions as a QUBO problem

Tesi di Laurea in Fisica Matematica

Relatore:	**Presentata da:**
Chiar.mo Prof.	**ALESSANDRO**
DAVIDE PASTORELLO	**BENEVENTI**

Correlatrice:
Dott.ssa
ANITA CAMILLINI

Anno Accademico 2024-2025

Abstract

Il molecular docking è una tecnica fondamentale nella biologia computazionale e nella progettazione di farmaci, volta a prevedere la configurazione di interazione più stabile tra un ligando e un recettore biologico. Negli ultimi anni, l'interesse verso l'integrazione di paradigmi computazionali avanzati – come il quantum annealing e il calcolo ad alte prestazioni – nei flussi di lavoro del docking è cresciuto notevolmente, con l'obiettivo di migliorarne accuratezza ed efficienza. Questa tesi contribuisce a tale ambito proponendo l'inclusione di un termine in grado di catturare le interazioni di π-stacking all'interno della funzione energetica globale (Hamiltoniana) utilizzata nei problemi di docking ligando-recettore. Le interazioni di π-stacking, che derivano da forze non covalenti tra sistemi aromatici, svolgono un ruolo cruciale nei processi di riconoscimento biomolecolare, ma sono spesso trascurate o trattate in modo semplificato nei modelli tradizionali. Nella prima parte del lavoro, sono state eseguite simulazioni di chimica computazionale su infrastrutture HPC per generare profili energetici dell'interazione tra anelli benzenici. Quindi, partendo da modelli QUBO precedenti, basati su rappresentazioni a grafo del docking molecolare, è stato formulato un nuovo termine hamiltoniano per rappresentare le interazioni di π-stacking, rispettando i vincoli imposti dal framework di ottimizzazione discreta. Particolare attenzione è stata rivolta all'integrazione di questo termine in modo compatibile con la natura discreta dei problemi di isomorfismo di sottografi. Infine, sono stati condotti calcoli mediante simulated annealing per validare il modello. I test, effettuati su sistemi in cui il benzene interagisce con residui aromatici come tirosina, fenilalanina e triptofano, hanno confermato che il nuovo termine influisce significativamente sulla configurazione di legame prevista. Questi risultati dimostrano la rilevanza dell'inclusione delle interazioni di π-stacking nei modelli di docking e aprono la strada a futuri miglioramenti nell'ottimizzazione molecolare guidata da principi fisici.

Abstract

Molecular docking is a fundamental technique in computational biology and drug design, aimed at predicting the most stable interaction configuration between a ligand and a biological receptor. In recent years, there has been growing interest in integrating advanced computational paradigms - such as quantum annealing and high-performance computing (HPC) - into docking workflows, with the goal of improving both accuracy and efficiency. This thesis contributes to this line of research by proposing the inclusion of a term capable of capturing π-stacking interactions within the global energy function (Hamiltonian) used in ligand–receptor docking problems. π-stacking interactions, which arise from non-covalent forces between aromatic systems, play a crucial role in biomolecular recognition processes but are often overlooked or treated simplistically in traditional docking models. In the first part of the work, computational chemistry simulations were performed on HPC infrastructures to generate detailed energy profiles of the interaction between benzene rings. Building on previous QUBO models based on graph representations of molecular docking, a new Hamiltonian term was formulated to account for π-stacking interactions, while respecting the constraints imposed by the discrete optimization framework. Particular attention was paid to integrating this term in a way that is compatible with the discrete nature of subgraph isomorphism problems. Finally, simulated annealing computations were carried out to validate the extended model. Tests were conducted on systems involving benzene interacting with aromatic residues such as tyrosine, phenylalanine, and tryptophan. The results confirmed that the newly introduced term significantly influences the predicted binding configuration, consistently guiding the ligand toward the correct spatial orientation. These findings demonstrate the relevance of incorporating π-stacking interactions into docking models and pave the way for future improvements in physically-informed molecular optimization.

Introduction

Molecular docking is a computational technique widely used in structural biology and drug discovery to predict the preferred orientation of two interacting molecules, typically a small ligand and a larger biological receptor, such as a protein. The primary goal of docking is to determine how these molecules fit together in three-dimensional space to form a stable complex, and to estimate the strength and nature of their interaction.

The ability to model and predict molecular binding with reasonable accuracy makes docking an invaluable tool in modern scientific research. In drug discovery, it allows for the virtual screening of vast chemical libraries, helping researchers identify promising candidate molecules with high affinity for a given biological target. This process significantly reduces both the time and cost associated with experimental screening, accelerating the early stages of drug development. Moreover, molecular docking contributes to a deeper understanding of the molecular mechanisms underlying biological functions and disease processes, providing essential insights for the rational design of new therapeutics.

Numerous studies in recent years have focused on improving both the methodologies and the accuracy of molecular docking. As the complexity of biological targets and chemical libraries increases, researchers are turning to increasingly sophisticated computational technologies to enhance docking performance and predictive power. In particular, advancements in High-Performance Computing (HPC) and the emerging field of Quantum Computing are opening new possibilities for tackling the combinatorial and energetic challenges inherent in molecular docking.

The work presented in this thesis is part of an ongoing research effort aimed at integrating these new computational paradigms into molecular docking workflows. This research line was initiated with the study *Molecular Docking via Weighted Subgraph Isomorphism on Quantum Annealers* [1], which approached the ligand-receptor docking problem by encoding geometric constraints into a Quadratic Unconstrained Binary Optimization (QUBO) formulation, making it suitable for solution via quantum annealing. This initial work focused on geometric compatibility between ligand and receptor.

Building upon this foundation, a subsequent study is currently focusing on the introduction of an additional physico-chemical constraint into the total system Hamiltonian, thereby enhancing the physical realism of the docking model. Both studies employed quantum and simulated annealing techniques to identify the equilibrium configuration of the ligand-receptor system, effectively minimizing the total energy as represented by the constructed Hamiltonian.

The primary objective of the present work is to further extend this research direction by incorporating a specific term into the physical Hamiltonian that explicitly accounts for π-stacking interactions. These interactions, which arise from non-covalent forces between aromatic rings, play a crucial role in many biomolecular recognition processes. By modeling π-stacking contributions within the Hamiltonian, the aim is to improve the energetic description of the docking problem and increase the accuracy of the predicted binding configurations.

Structure of the Thesis

First and Second Chapter: Preliminaries

The first two chapters of this thesis provide the necessary theoretical background to contextualize the research, simulations, and modeling work presented in Chapters 3, 4 and 5. This preliminary section is designed to provide the reader with a clear understanding of the computational and physical principles that underpin the subsequent developments.

The first chapter begins by introducing the concepts of Quadratic Unconstrained Binary Optimization (QUBO), simulated annealing, and quantum annealing. These methodologies are described in detail, with particular emphasis on their practical applications, their relevance to optimization problems, and the interconnections between them. Leonardo supercomputer is also introduced.

Following this, in the second chapter, a comprehensive overview of molecular docking is presented. This section explains the fundamental principles of docking and its significance in computational chemistry and drug discovery. After briefly reviewing preliminary studies in the field, special attention is devoted to the key reference work underpinning this thesis, namely *Molecular Docking via Weighted Subgraph Isomorphism on Quantum Annealers* [1]. Additionally, an overview is provided on ongoing research related to the physico-chemical Hamiltonian, as this thesis aims to extend and further develop that line of inquiry.

The chapter then proceeds with an in-depth discussion of π-stacking interactions, starting with the foundational Hunter–Sanders model and moving on to more recent developments in the theoretical understanding of these non-covalent forces. This section highlights the role of π-stacking in biomolecular recognition and its relevance to the present study.

Finally, the reader is introduced to the field of computational chemistry. Particular attention is given to the theoretical foundations and practical applications of Density Functional Theory (DFT) and *ab initio* methods, both of which have been essential for generating the reference data used in this work.

Third Chapter: HPC simulations

The second chapter focuses on the computational simulations performed in a High-Performance Computing (HPC) environment using the Leonardo supercomputer at CINECA. An extensive set of simulations was conducted to obtain a comprehensive and detailed characterization of the π-stacking interaction. Although this interaction has been widely studied from a chemical perspective to identify its dominant energetic contributions, no closed-form physical equation currently exists to describe it quantitatively.

To address this gap, it was necessary to explore the interaction strength across various spatial configurations and relative orientations between aromatic rings, as such structures are typically those that display this characteristic behavior. In particular, simulations have been conducted using benzene rings, the simplest form of aromatic structures. This systematic exploration allowed us to gather the data required for developing an effective mathematical model of the interaction.

In section 3.1 and 3.2 we present the computational codes and simulation protocols used throughout the study, including those implemented with Quantum ESPRESSO and Q-Chem. This is followed by a detailed presentation of the simulation results (3.3), highlighting the key features of the energy landscapes obtained.

Finally, Section 3.4 is dedicated to the polynomial fitting of the calculated energy surfaces. These fits were developed to provide continuous functional representations of the interaction energy, which will later serve as input for the Hamiltonian modeling stage of the project.

Fourth Chapter: Mathematical Modelling

The third chapter (3) represents the core of this work. Here, we address the mathematical modeling required to incorporate the π-stacking contribution into the physical Hamiltonian introduced in the previously discussed study.

This task presented several challenges. The original model for describing ligand-receptor interactions [1] was primarily designed to account for atomistic interactions, such as electrostatics and van der Waals forces, where each atom is represented as a node within a graph-based structure. While this graph formalism is well-suited for modeling interatomic forces, it is not inherently designed to capture interactions between larger molecular fragments, such as aromatic rings.

The central difficulty arose from the geometric constraints associated with defining the position and orientation of an aromatic ring in three-dimensional space. Accurately positioning a planar aromatic ring typically requires fixing three of its constituent atoms. However, the QUBO formulation adopted in the existing model allows for direct control over only two atoms at a time, as the variables in the optimization process correspond to discrete positional choices for individual atoms.

Fortunately, the results obtained from the computational simulations provided a solution to this limitation. The energy landscapes revealed that one degree of freedom in the movement of the aromatic rings could be reasonably neglected without compromising the accuracy of the model. This insight enabled the development of a modified Hamiltonian that incorporates the π-stacking contribution while remaining compatible with the existing framework. The resulting formulation successfully integrates into the original optimization model, allowing for the inclusion of physically meaningful π-stacking interactions within the ligand-receptor docking process.

Fifth Chapter: Annealing and Results

In Chapter 4, we carried out a series of simulated annealing computations to test the validity and effectiveness of the model introduced in Chapter 3. Specifically, our goal was to evaluate the impact of the π-stacking term incorporated into the global Hamiltonian used to describe the total energy of the system. Through these calculations, we demonstrated that the inclusion of the π-stacking contribution is essential for correctly predicting the spatial configuration of a benzene fragment interacting with amino acids such as tyrosine, phenylalanine, and tryptophan. By exploring a range of configurations - altering both the pocketgrid and the spatial grid structure - we obtained consistent and significant results that reinforce the central role of π-stacking in determining the optimal binding arrangement. These findings provide a robust computational foundation for modeling non-covalent interactions involving aromatic systems within protein-ligand environments.

Contents

Chapter 1

Computational Setup: Leonardo and Annealing

1.1 Introduction to QUBO, Simulated and Quantum Annealing

Quadratic Unconstrained Binary Optimization (QUBO) is a mathematical framework for modeling combinatorial optimization problems whose variables take binary values $x_i \in \{0, 1\}$, and whose objective function is a quadratic form in these variables. Formally, given a symmetric real-valued matrix $Q \in \mathbb{R}^{n \times n}$, the goal is to find a binary vector $x \in \{0, 1\}^n$ that minimizes the cost function

$$f_Q(x) = x^\intercal Q x.$$

Due to the exponential growth of the solution space with the number of variables, solving QUBO problems exactly becomes computationally intractable for large instances [2]. This has led to the development of approximate algorithms capable of finding near-optimal solutions efficiently.

In particular, QUBO problems can be effectively addressed using both *simulated annealing* and *quantum annealing* [2], two optimization techniques inspired by thermodynamic and quantum physical principles, respectively.

Simulated annealing, which is introduced in Section 1.1.2, is a classical metaheuristic algorithm that explores the solution space probabilistically, allowing for both downhill and uphill moves according to a temperature-dependent acceptance rule. This probabilistic flexibility enables the algorithm to escape local minima and makes it particularly well suited to minimizing the energy landscapes defined by QUBO objectives.

Quantum annealing, introduced in Section 1.1.3, relies on quantum mechanical effects such as tunneling and superposition to explore the solution space. It operates on the same class of problems, and in fact, most physical implementations of quantum annealers (e.g., D-Wave systems) natively solve problems formulated as QUBO or equivalently Ising models.

In the following sections, we first present the definition and structure of a QUBO problem, then we will introduce the simulated annealing algorithm in detail, and subsequently explore its quantum analogue.

1.1.1 Quadratic Unconstrained Binary Optimization (QUBO)

The Quadratic Unconstrained Binary Optimization (QUBO) is a central mathematical framework used to express a wide range of discrete optimization problems across domains such as finance, logistics, and machine learning [3].

Its relevance extends beyond classical computation: QUBO problems are directly translatable into the Ising model, making them highly suitable for solution methods based on quantum annealing and other adiabatic quantum algorithms.

Let $\mathbb{B} = \{0, 1\}$ denote the binary set, and $\mathbb{B}^n$ the space of binary vectors of length $n \in \mathbb{N}$. Given a real symmetric matrix $Q \in \mathbb{R}^{n \times n}$, the standard QUBO objective function takes the form

$$f_Q(x) = x^\intercal Q x = \sum_{i=1}^{n} \sum_{j=1}^{n} Q_{ij} x_i x_j,$$

where $x \in \mathbb{B}^n$ is the binary decision vector.

An alternative but equivalent formulation separates the purely quadratic terms from the linear contributions:

$$f_{Q',q}(x) = x^\intercal Q' x + q^\intercal x,$$

with $Q' \in \mathbb{R}^{n \times n}$ symmetric and $q \in \mathbb{R}^n$. The two formulations are equivalent by setting $Q = Q' + \mathrm{diag}(q)$, taking advantage of the identity $x_i^2 = x_i$ for binary variables.

The optimization goal is to find $x^* \in \mathbb{B}^n$ that minimizes the objective:

$$f_Q(x^*) \leq f_Q(x) \quad \forall x \in \mathbb{B}^n.$$

Since the number of candidate solutions is exponential in n, i.e., 2^n, QUBO problems are generally computationally intensive for large systems.

Nevertheless, specific QUBO instances admit efficient solutions under structural constraints. For example if $Q_{ij} > 0$ for all i, j, the global minimum is trivially $x^* = 0$, if $Q_{ij} < 0$ for all i, j, the minimum is $x^* = 1$ and if Q is diagonal, each variable decouples and can be optimized in linear time.

Overall, QUBO serves as a powerful and versatile framework for encoding discrete optimization tasks, and is central to the development of quantum-inspired and quantum-native algorithms.

Equivalence with the Ising Model

QUBO formulations are the matematical equivalent of the classical Ising model from statistical mechanics, which describes systems of spins under pairwise interactions and external magnetic fields.

In the Ising formulation, the energy of a spin configuration $\sigma \in \{-1, +1\}^n$ is expressed as

$$H(\sigma) = \sigma^\intercal J \sigma + h^\intercal \sigma = \sum_{i,j} J_{ij} \sigma_i \sigma_j + \sum_i h_i \sigma_i,$$

where $J \in \mathbb{R}^{n \times n}$ encodes pairwise interactions (symmetric), and $h \in \mathbb{R}^n$ represents the external field acting on each spin. To map an Ising model to a QUBO instance, we apply the transformation

$$\sigma = 1 - 2x,$$

with $x \in \{0,1\}^n$. Substituting into $H(\sigma)$, we obtain:

$$\begin{aligned} H(x) &= (1-2x)^\intercal J(1-2x) + h^\intercal(1-2x) \\ &= 4x^\intercal Jx - 4(J1)^\intercal x + \text{const.} \end{aligned}$$

The constant term can be omitted from the optimization. Therefore, the corresponding QUBO matrix is $Q = 4J$, and the linear term is $q = -4J1+2h$, yielding a QUBO problem equivalent to the original Ising model.

The reverse mapping is also straightforward. Given a QUBO matrix Q, we can define the Ising couplings and fields as:

$$J = \frac{Q}{4}, \qquad h = -\frac{Q1 + Q^\intercal 1}{4}.$$

The energy offset is

$$E_0 = \frac{1^\intercal Q1}{4}.$$

This two-way mapping allows one to switch between the QUBO and Ising representations depending on the solver or physical model being used.

1.1.2 Simulated Annealing

Simulated annealing (SA) is a probabilistic metaheuristic technique used to approximate the global optimum of an objective function, particularly in optimization problems characterized by a large and rugged search space with many local optima. When the goal is to obtain a near-global optimum within a limited computational budget, simulated annealing often outperforms deterministic methods like gradient descent or branch and bound. The method was formalized by Kirkpatrick, Gelatt, and Vecchi in 1983 [4], building upon ideas originating from the Metropolis–Hastings algorithm [5], a Monte Carlo technique introduced in 1953 for simulating thermodynamic systems. Related ideas had also appeared earlier in the work of Pincus [6] and Khachaturyan et al. [7, 8].

The algorithm draws its name from the physical process of *annealing* in metallurgy, where a material is heated and then slowly cooled to reduce its internal defects and minimize its thermodynamic free energy. This analogy is central to the design of the algorithm: at high "temperatures," the system can accept worse solutions, thus exploring a wide region of the search space; as the temperature decreases, the algorithm becomes increasingly selective, favoring low-energy configurations and converging toward an optimum.

Simulated annealing has proven effective in situations where exact algorithms are computationally infeasible or prone to becoming trapped in suboptimal regions. Although convergence to the global minimum is not guaranteed in practice, SA often produces high-quality solutions suitable for real-world applications.

The algorithm

The algorithm begins with an initial state s and iteratively explores the solution space. At each step, it generates a neighboring state s^* through a local perturbation of s, computes its energy $E(s^*)$, and decides whether to move to s^* based on a probabilistic acceptance criterion. This process is repeated until a stopping condition is met—typically when a maximum number of iterations is reached or when no significant improvement is observed.

The decision to accept a new state s_{new} depends on the associated energies $e = E(s)$ and $e_{\text{new}} = E(s_{\text{new}})$, and on the current temperature T. The most commonly used acceptance probability is given by the Metropolis criterion:

$$P(e, e_{\text{new}}, T) = \begin{cases} 1, & \text{if } e_{\text{new}} < e, \\ \exp\left(-\dfrac{e_{\text{new}} - e}{T}\right), & \text{if } e_{\text{new}} \geq e. \end{cases}$$

This rule ensures that all energy-decreasing moves are accepted, while energy-increasing moves are accepted with a probability that decays exponentially with the increase in energy and the reduction in temperature. As a result, at high temperatures, the algorithm explores widely, accepting worse solutions with high probability, while at low temperatures it becomes more selective and converges locally.

The temperature T plays a critical role in controlling the balance between exploration and exploitation. When T is large, the system is tolerant to energy increases, allowing it to escape local minima and explore distant regions of the state space. As T decreases, the algorithm becomes increasingly conservative, accepting only moves that result in lower or slightly higher energy. In the limit $T \to 0$, SA behaves like a greedy local search algorithm.

The evolution of T is governed by the *annealing schedule*, which defines how the temperature is reduced over time. A common schedule is geometric:

$$T_k = \alpha^k T_0,$$

where $\alpha \in (0, 1)$ is the cooling rate, T_0 is the initial temperature, and k is the iteration count. Alternative schedules, such as logarithmic cooling, may be used to meet theoretical convergence criteria, though they are often too slow for practical use.

In contrast to greedy heuristics like hill climbing—which always move to the best available neighbor and terminate at local optima—simulated annealing can accept worse solutions and therefore escape local traps. This makes it especially effective in rugged or multimodal landscapes, where purely local strategies are insufficient. The ability to probabilistically accept uphill moves distinguishes SA from traditional local search and gives it global search capabilities, provided the temperature decreases slowly enough and the search is allowed to run sufficiently long.

Simulated annealing has been shown to converge to the global optimum with probability one under specific conditions. In particular, for a finite state space and a logarithmic cooling schedule, it holds that:

$$\lim_{k \to \infty} \mathbb{P}(s_k \in \mathcal{S}^*) = 1,$$

where $\mathcal{S}^*$ denotes the set of global optima and s_k is the state at iteration k. However, the rate of convergence is extremely slow, and the time required to guarantee a high success probability is often impractical. In applied settings, simulated annealing is valued for its ability to find high-quality solutions efficiently, rather than for its asymptotic guarantees.

1.1.3 Quantum Annealing

Quantum annealing (QA) is a metaheuristic method inspired by quantum mechanics and developed to solve optimization problems by exploiting the quantum properties of physical systems. It shares important conceptual features with adiabatic quantum computing (AQC), yet the two paradigms are not equivalent. While AQC assumes a perfectly isolated system evolving unitarily according to the Schrödinger equation, QA considers an open quantum system interacting with its environment, where decoherence and dissipation play a central role. The following discussion is based on the treatment presented in [9]. For further details and a deeper understanding, the reader is referred to the following references: [10], [11], [12].

Adiabatic Quantum Computing

Adiabatic quantum computing is grounded in the adiabatic theorem of quantum mechanics, which states that a quantum system initially prepared in the ground state $|\psi_0\rangle$ of a Hamiltonian H_I will remain close to the ground state of a slowly evolving Hamiltonian $H(t)$, provided that the evolution is sufficiently slow and a nonzero spectral gap is maintained throughout the evolution. The system evolves according to the Schrödinger equation:

$$i\frac{d}{dt}|\psi(t)\rangle = H(t)|\psi(t)\rangle, \quad t \in [0,T], \tag{1.1}$$

with initial condition $|\psi(0)\rangle = |\psi_0\rangle$.

By reparametrizing time with $s = t/T \in [0,1]$, the Hamiltonian can be written as $\widetilde{H}(s) = H(Ts)$. Let $E_0(s)$ be the lowest eigenvalue of $\widetilde{H}(s)$ and $|0,s\rangle$ the corresponding eigenstate. Let us assume the non-degeneracy of the initial ground state, $\dim \mathcal{H}_{E_0(0)} = 1$, where $\mathcal{H}_{E_0(0)}$ is the eigenspace corresponding to $E_0(0)$. If the spectral gap $\lambda(s) = E_1(s) - E_0(s) > 0$ for all $s \in [0, 1]$, then the adiabatic theorem ensures:

$$\lim_{T \to \infty} |\langle 0, 1 | \psi(T) \rangle| = 1. \tag{1.2}$$

This means the final state $|\psi(T)\rangle$ is with high probability the ground state of the final Hamiltonian $H_P = H(T)$.

The required runtime T to ensure adiabaticity can be estimated as:

$$T \gg \frac{\max_s \left\| \frac{d\widetilde{H}(s)}{ds} \right\|_{op}}{\min_s \lambda(s)^2}, \tag{1.3}$$

where $\| \cdot \|_{op}$ denotes the operator norm. The standard adiabatic quantum computing protocol consists of:

- Preparing the system in the ground state of a simple Hamiltonian H_I.
- Slowly evolving the system under a Hamiltonian $H(t) = (1 - s(t))H_I + s(t)H_P$ with $s(0) = 0$ and $s(T) = 1$.
- Measuring the final state to obtain the ground state of the problem Hamiltonian H_P.

Quantum Annealing

In quantum annealing, the problem to solve is encoded in the ground state of a Hamiltonian H_P acting on the Hilbert space $\mathcal{H}$ of n qubits. The system is driven toward this ground state via a time-dependent Hamiltonian:

$$H(t) = \Gamma(t)H_D + H_P, \tag{1.4}$$

where H_D is the transverse-field or "disordering" Hamiltonian, which induces quantum fluctuations, and $\Gamma(t)$ is a decreasing function of time.

Typically, the initial Hamiltonian H_D is chosen so that its ground state is a simple product state that can be easily prepared experimentally. A common and practical choice is:

$$H_D = -\sum_{i=1}^{N} \Delta_i \sigma_i^x,$$

where $\Delta_i > 0$ are real coefficients and σ_i^x denotes the Pauli-X operator acting on the i-th qubit. This Hamiltonian corresponds to a transverse magnetic field that induces quantum transitions (bit flips) between the computational basis states $\{|0\rangle, |1\rangle\}$ of each qubit. This term introduces quantum fluctuations that enable the system to:

- transition between different configurations by flipping the states of individual qubits;
- induce quantum tunneling between distinct local minima in the energy landscape.

When the transverse field is strong, the system can freely explore a wide range of configurations due to enhanced quantum fluctuations. As the field is gradually reduced, the influence of the problem Hamiltonian becomes dominant, and the system is guided toward the optimal solution encoded in its ground state.

The ground state of H_D in this case is a uniform superposition over all possible computational basis states, given by:

$$|+\cdots+\rangle = \frac{1}{\sqrt{2^N}} \sum_{z\in\{0,1\}^N} |z\rangle,$$

where each qubit is in the eigenstate $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ of σ^x. This state is highly symmetric and can be efficiently prepared in physical quantum hardware by applying Hadamard gates to each qubit initialized in the state $|0\rangle$, where the Hadamard gate is a single-qubit gate defined in matrix form with respect to the computational basis $\{|0\rangle, |1\rangle\}$ as follows:

$$H := \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Choosing such an H_D ensures that the system starts in a delocalized quantum state, from which the adiabatic or annealing evolution can begin to guide it toward the ground state of the problem Hamiltonian H_P. The problem Hamiltonian has the form:

$$H_P = \sum_{i \in V} \theta_i \sigma_i^z + \sum_{(i,j) \in E} \theta_{ij} \sigma_i^z \sigma_j^z, \tag{1.5}$$

where $\sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is the Pauli Z matrix acting on qubit i, and θ_i, θ_{ij} are parameters representing local fields and couplings.

The eigenvalues of H_P correspond to the cost function of the classical Ising model:

$$E(z) = \sum_{i \in V} \theta_i z_i + \sum_{(i,j) \in E} \theta_{ij} z_i z_j, \quad z \in \{-1, 1\}^{|V|}, \tag{1.6}$$

and the goal is to find the configuration z^* that minimizes $E(z)$:

$$z^* = \arg \min_{z \in \{-1,1\}^{|V|}} E(z). \tag{1.7}$$

During the annealing process, the system is initialized with $\Gamma(0)$ large (dominated by H_D), and as $\Gamma(t)$ decreases, the system is guided towards the ground state of H_P, ideally ending in the optimal configuration. The binary variables z_i are physically realized by the final measurement outcomes of the qubits.

Effectively applying quantum annealing requires translating the optimization objective into the parameters θ_i and θ_{ij} of the Ising model, which—as shown in Section 1.1.1—is equivalent to the corresponding QUBO formulation.

Simulated and Quantum Annealing with D-Wave

For the simulated annealing computations presented in Chapter 5, we employed the `neal.SimulatedAnnealingSampler` [13] from the `neal` Python package. `neal` is part of the Ocean software development toolkit developed by D-Wave Systems, which provides a suite of open-source tools designed to formulate, manipulate, and solve optimization problems -particularly those expressed as QUBO models - on both quantum and classical hardware. The Ocean ecosystem includes components for problem embedding, solver interfacing, and result analysis, and is widely used in quantum annealing research and applications.

The `SimulatedAnnealingSampler` in `neal` is a classical solver that mimics the physical process of annealing by iteratively searching for a configuration of variables that minimizes the energy of a system. In our experiments, this sampler utilizes a *geometric beta schedule*, which governs the progression of the inverse temperature parameter β throughout the annealing process.

A geometric beta schedule increases β multiplicatively at each step, following the rule $\beta_{k+1} = r \cdot \beta_k$, where $r > 1$ is a constant scaling factor. This approach ensures that the temperature decreases exponentially, allowing the system to explore the solution space broadly at high temperatures and gradually converge toward low-energy configurations as the temperature cools. This type of scheduling is particularly effective in avoiding local minima and improving convergence toward the global optimum in complex energy landscapes.

D-Wave Systems is also among the most prominent and commercially available implementations of quantum annealing hardware. Unlike gate-based quantum computers, which operate using quantum logic gates and circuit models, D-Wave machines are designed specifically to perform quantum annealing for solving combinatorial optimization problems. These devices consist of thousands of superconducting qubits arranged according to a fixed hardware topology with limited connectivity. As a consequence, translating a logical optimization problem onto the D-Wave hardware requires an additional step known as embedding, which maps the problem variables onto the available physical qubits while respecting their connectivity constraints.

The process of embedding is a crucial step in solving optimization problems on a D-Wave quantum annealer. Given the hardware constraints of the machine, embedding is necessary to map the logical structure of the problem onto the physical layout of the qubits. A QUBO problem, or equivalently its Ising formulation, can be represented as a graph where each node corresponds to a binary variable and each edge represents a quadratic interaction between variables. This logical graph reflects the connectivity required by the problem. However, the physical qubits in a D-Wave machine are arranged according to a specific hardware topology, such as Chimera, Pegasus, or Zephyr [14]. These topologies are sparse by design, meaning that not all pairs of qubits are directly connected. As a result, the embedding process involves finding a mapping between the logical variables and the available physical qubits, taking into account the connectivity constraints of the device. When a logical variable requires connections to multiple other variables, and these connections cannot be satisfied by a single physical qubit due to limited connectivity, the embedding algorithm assigns the logical variable to a chain of multiple physical qubits. These qubits are strongly coupled to enforce that they behave coherently as a single logical entity during the annealing process. This allows the hardware to simulate higher connectivity than what is physically available.

To illustrate this concept, consider the following simple QUBO problem defined over two binary variables:

$$Q = \{(0,0) : -1,\ (1,1) : -1,\ (0,1) : 2\}.$$

The corresponding objective function is:

$$E(x_0, x_1) = -x_0 - x_1 + 2x_0x_1,$$

where x_0 and x_1 are binary variables. This problem can be represented as a logical graph with two nodes connected by a single edge, reflecting the quadratic interaction between x_0 and x_1. If the physical topology of the D-Wave machine provides two qubits that are directly connected, the embedding for this problem is straightforward: each logical variable is mapped to a distinct physical qubit, and the coupling between them represents the interaction term in the QUBO.

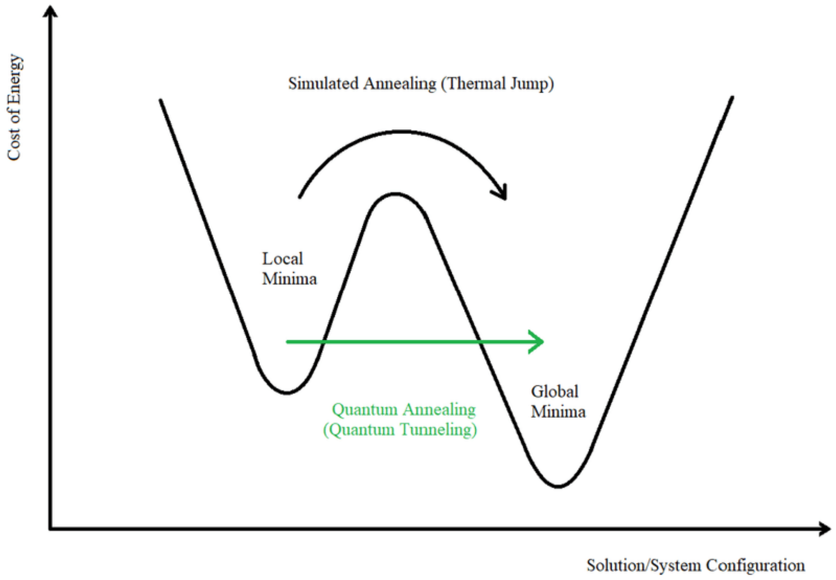


Figure 1.1: Illustration of the difference between simulated annealing and quantum annealing in escaping local minima. Simulated annealing relies on thermal jumps to overcome energy barriers, while quantum annealing exploits quantum tunneling. Image adapted from [15].

In more complex scenarios, where the required logical connectivity exceeds the available physical connections, the embedding process becomes non-trivial. The embedding algorithm then identifies suitable chains of physical qubits to represent each logical variable while preserving the problem's connectivity. The quality of the embedding can significantly influence the success of the annealing process, as longer chains are more susceptible to errors such as chain breaks. Consequently, embedding optimization remains a key aspect of practical quantum annealing on current hardware architectures.

Analogies between Quantum and Simulated Annealing

Quantum annealing can be meaningfully compared to simulated annealing, with an important analogy between their respective control parameters: the *temperature* in simulated annealing plays a role analogous to the *transverse field strength* in quantum annealing.

In simulated annealing, the temperature governs the probability of accepting transitions to higher-energy configurations, allowing the system to escape local minima by thermal activation. As the temperature is gradually reduced, the system becomes increasingly selective, eventually converging to a low-energy state.

In quantum annealing, quantum tunneling - induced by the transverse field - enables the system to explore the energy landscape by coherently modifying the amplitudes of many states in parallel. The strength of the transverse field controls the extent of these quantum fluctuations and is gradually decreased to guide the system toward the ground state of the problem Hamiltonian.

Both methods aim to avoid local minima and reach the global optimum, but use fundamentally different physical mechanisms: thermal jumps in simulated annealing versus quantum tunneling in quantum annealing. Analytical [12] and numerical [16] studies suggest that, under specific conditions, quantum annealing can outperform simulated annealing, especially in problems with tall and narrow energy barriers, where quantum tunneling is more efficient than thermal hopping.

1.2 HPC and Leonardo Supercomputer

High-Performance Computing (HPC) refers to the use of supercomputers and parallel processing techniques to solve complex computational problems that require significant processing power and memory resources. HPC systems are characterized by their ability to perform trillions of calculations per second, enabling the simulation, modeling, and analysis of large-scale scientific, engineering, and industrial problems that would be intractable using conventional computing systems.

A prominent example of HPC infrastructure in Europe is the Leonardo supercomputer, located at CINECA in Bologna, Italy. Leonardo is part of the EuroHPC Joint Undertaking initiative, aimed at developing a pan-European supercomputing ecosystem. Operational since 2022, Leonardo is classified as a pre-exascale system, capable of delivering peak performance exceeding 250 petaflops (quadrillions of floating-point operations per second).

Leonardo is built on a heterogeneous architecture that combines traditional CPU-based nodes with highly parallel GPU-accelerated nodes, making it particularly suited for a wide range of applications including artificial intelligence, big data analytics, and large-scale scientific simulations. Its high-performance interconnect and large storage capacity allow efficient handling of massive datasets and enable parallel execution of computationally intensive workloads.

Within the context of computational chemistry and quantum simulations, HPC systems like Leonardo provide the necessary computational resources to perform ab initio calculations, molecular dynamics simulations, and quantum Monte Carlo methods on systems of realistic size and complexity.

Chapter 2

Background: π-Stacking, Quantum Molecular Docking & Computational Approaches

2.1 An overview on Quantum Molecular Docking

Molecular docking is a key computational tool in structural biology and drug design, aimed at predicting the binding configuration of two interacting molecules, typically a small ligand and a larger biological target such as a protein. By simulating the molecular recognition process, docking enables the identification of energetically favorable binding poses and provides insight into the strength and nature of molecular interactions. It plays a crucial role in the rational design of therapeutic compounds and in the interpretation of biochemical mechanisms at the molecular level.

In this chapter, we briefly present the general framework of molecular docking and its formulation as an optimization problem. We then introduce in details a first work which explores molecular docking from a quantum computational perspective. *Molecular Docking via Weighted Subgraph Iso-*

morphism on Quantum Annealers [1] indeed models the docking process as a purely geometric problem, leveraging weighted subgraph matching techniques. A second approch, currentely still being investigated, extends this framework by incorporating physically motivated interaction terms into the definition of the system's Hamiltonian. Both approaches are designed to be compatible with current quantum annealing hardware and provide innovative formulations of molecular docking within the discrete optimization paradigm.

2.1.1 Classical Molecular Docking

Molecular docking is a computational technique used in molecular modelling to predict the preferred orientation of two molecules when they interact to form a stable complex. In most applications, the goal is to determine how a small molecule ligand binds to a macromolecular target, such as a protein, in a way that mimics biological recognition processes. Once a plausible binding configuration is identified, the strength of the interaction - commonly referred to as binding affinity - can be estimated using appropriate scoring functions.

This methodology is widely employed in structure-based drug design, where it provides valuable insights into the binding behavior of candidate compounds. Accurate prediction of ligand-target interactions is essential not only for the rational design and optimization of new drug molecules, but also for understanding key biochemical and pharmacological mechanisms. Molecular docking is a cornerstone of virtual screening workflows, enabling the evaluation of large chemical libraries to identify promising molecules for experimental validation. Furthermore, docking simulations serve as an effective starting point for lead compound optimization and mechanistic studies.

From a biological perspective, docking plays a central role in modeling the association between biologically relevant molecules, including proteins, peptides, nucleic acids, carbohydrates, and lipids. The spatial arrangement adopted by the interacting species can significantly influence the functional outcome of the interaction. For example, different binding conformations may elicit agonistic or antagonistic responses in the target system. As such, molecular docking not only assesses interaction strength but also contributes to the interpretation of functional effects driven by molecular recognition.

Conceptually, molecular docking can be formulated as a geometrical and energetic optimization problem. A classical analogy is the "lock-and-key" model, in which the protein represents the lock and the ligand the key; the objective is to identify the orientation in which the ligand best fits the binding site. However, because both ligands and receptors often exhibit conformational flexibility, a more accurate representation is the "hand-in-glove" model. In this scenario, both molecules undergo conformational changes to achieve a better fit, a process known as *induced fit*. This mutual adaptation stabilizes the resulting protein-ligand complex.

Docking algorithms attempt to simulate this molecular recognition process by exploring a large number of potential configurations and evaluating them according to energy-based criteria. The ultimate goal is to identify the relative orientation and conformation of the interacting molecules that minimizes the system's free energy. This typically involves heuristic or stochastic search strategies, which allow the efficient sampling of the configuration space and the selection of energetically favourable binding poses.

In recent years, advances in quantum computing have opened new avenues for addressing molecular docking through quantum-inspired methodologies. Several studies have proposed quantum approaches to reformulate or support various aspects of the docking process. Banchi et al. [17] introduced the use of Gaussian Boson Sampling to tackle a rigid-body formulation of docking by mapping the problem to the search for maximum cliques in a graph representation of the molecular structures. In a different context, Mensa et al. [18] applied a quantum-enhanced machine learning strategy - Quantum Kernel Estimation - to the problem of Ligand-Based Virtual Screening (LB-VS), a technique closely related to docking.

More recently, Mato et al. [19] investigated the use of Quantum Annealing techniques (see Section 1.1.3) to support docking by focusing on the ligand expansion phase. Their study addressed the problem of identifying torsional configurations of the ligand that maximize its molecular volume, or equivalently, that maximize the internal atomic distances, thus promoting favorable spatial arrangements for subsequent binding.

In the following section, we present the work *Molecular Docking via Weighted Subgraph Isomorphism on Quantum Annealers* [1], which addresses the entire Search and Construction (SC) phase of molecular docking by reformulating the problem in a manner more compatible with quantum annealing. Instead of relying on conventional methodologies based on rigid-body roto-translations and fragment rotations, this approach introduces a novel problem representation that aligns naturally with the structure of a QUBO problem.

Subsequently, we discuss the ongoing research building upon the framework proposed in [1]. This later work estends the original formulation by incorporating constraints derived from the physico-chemical interactions between molecules, thereby going beyond purely geometric considerations and enhancing the physical realism of the docking process.

2.1.2 Molecular Docking via Weighted Subgraph Isomorphism on Quantum Annealers

The approach proposed in [1] introduces a fully geometric formulation of the molecular docking problem based on the identification of a set of representative docking points inside the binding pocket. These points, referred to as pocket probes, define the active region of the pocket and are selected based on its shape and volume using established tools such as CAVIAR, PASS, and POCASA [20]. The resulting probe points serve as the nodes of a weighted spatial grid, providing a discrete approximation of the 3D cavity. Edges between points are weighted by their Euclidean distance, yielding a graph representation $G_{\text{grid}} = \{v, e_{u,v}, w_{u,v}\}$ where each node v corresponds to a docking point, the terms $e_{u,v}$ represent the edges between one node and another, and each edge weight $w_{u,v}$ is given by $\|u - v\|$.

Ligands are encoded as weighted graphs $G_{\text{mol}} = \{i, e_{i,j}, w_{i,j}\}$, where each node i corresponds to a ligand atom, the terms $e_{i,j}$ represent the edges between one node and another, and, as before, each edge weight $w_{i,j}$ is given by $\|i - j\|$. Ligands are designed to preserve both chemical and geometric information. This includes atom connectivity, rotatable bonds, bond lengths, and fixed bond angles. To construct these graphs, a pre-processing phase is applied which performs the following simplifications:

- Removal of terminal hydrogens: Peripheral hydrogen atoms are excluded due to their minimal impact on the global shape of the ligand.
- Identification of rotatable bonds and fragments: The ligand is partitioned into rigid fragments determined by the location of rotatable bonds.
- Fragment simplification: Structural simplifications were applied to the ligand fragments to reduce complexity while preserving key geometric features.

Following this, the ligand graph captures a simplified yet informative molecular structure, where nodes represent atoms or fragments, and weights encode geometric constraints such as bond lengths and angular relationships.

The docking task is then reformulated as a weighted subgraph isomorphism problem, where the objective is to find an optimal mapping between the ligand graph G_{mol} and the spatial grid G_{grid}. To model this as a QUBO problem, binary variables $x_{i,i'} \in \{0, 1\}$ are introduced to represent whether a node $i \in G_{\text{mol}}$ is mapped to a node $i' \in G_{\text{grid}}$. The mapping is constrained to be injective through the penalty term:

$$H_1 = \sum_{i} \left(1 - \sum_{i'} x_{i,i'} \right)^2 .$$

To enforce structural consistency between the two graphs, a second term penalizes invalid edge mappings, i.e., if an edge exists between nodes i and j

in the ligand graph, but their corresponding images i' and j' are not connected in the grid:

$$H_2 = \sum_{(i,j)\in G_{\text{mol}}} \sum_{(i',j')\notin G_{\text{grid}}} x_{i,i'} x_{j,j'}.$$

Together, these define the structural feasibility component:

$$H_{\text{iso}} = H_1 + H_2.$$

To preserve geometric accuracy, a third term H_{opt} penalizes discrepancies between ligand edge weights and their mapped counterparts in the spatial grid:

$$H_{\text{opt}} = \sum_{(i,j)\in G_{\text{mol}}} \sum_{(i',j')\in G_{\text{grid}}} (w_{i,j} - w_{i',j'})^2 \, x_{i,i'} \, x_{j,j'}.$$

The total QUBO Hamiltonian becomes:

$$H_{\text{QUBO}} = AH_{\text{iso}} + BH_{\text{opt}},$$

where $A \gg B$ ensures that hard constraints are always prioritized. Without loss of generality, B can be set to 1 and A treated as the main tuning parameter.

This formulation is particularly well suited for implementation on quantum annealers. One of the key strengths of this method lies in its natural compatibility with the QUBO formalism. The study then evaluates the performance of this geometric approach on quantum hardware and compares it with classical techniques such as simulated annealing, highlighting both the current capabilities and limitations of quantum annealing for practical molecular docking tasks.

2.1.3 Adding Physicochemical Constraints into the Geometric Hamiltonian

While the previous article primarily focused on optimizing the geometric compatibility between the ligand and the binding pocket, the authors of this work propose an extended formulation that incorporates both geometric and physicochemical properties. In contrast to approaches that rely on pharmacophoric points [21]—which serve only as indirect representations of key ligand-protein interactions [22]—this method preserves a stronger dependence on the ligand's full structure.

Building upon the geometric Hamiltonian formulation proposed in [1], this new research extends the model by introducing physicochemical interaction terms such as Coulomb and van der Waals. These terms aim to enhance the biological realism of the docking formulation and are later embedded into a QUBO Hamiltonian.

To incorporate such interactions, the protein is no longer treated solely as a static surface defining the pocket. Instead, its atomic structure is used to compute the influence each protein atom exerts on the surrounding 3D space. These effects are precomputed and encoded in the weights of the spatial grid G_{grid}.

- *Coulomb Interaction*: The electrostatic potential generated by the protein is modeled as a sum over Coulomb potentials. For a given point $j \in G_{\text{grid}}$, the potential is computed as:

  $$V_j = \frac{1}{4\pi\varepsilon} \sum_k \frac{q_k}{r_{jk}},$$

  where q_k is the charge of protein atom k, r_{jk} is its distance from point j, and ε is the dielectric constant. This scalar potential is used to assign an electrostatic weight to each grid node j, while ligand atoms are annotated with their corresponding charges. The electrostatic contribution of each atom–grid mapping is then simply the product of these quantities.

- *van der Waals Interaction*: The van der Waals contribution is modeled using the Lennard-Jones potential:

  $$U_{LJ}^{ik} = \sum_k \varepsilon_{ik} \left[\left(\frac{R_{\text{min},ik}}{r_{ik}} \right)^{12} - 2 \left(\frac{R_{\text{min},ik}}{r_{ik}} \right)^{6} \right],$$

> where r_{ik} is the distance between atoms i (ligand) and k (protein), ε_{ik} the well depth, and $R_{\min,ik}$ the distance at which the potential reaches its minimum. The mixing parameters are calculated using the Lorentz-Berthelot rules: $\varepsilon_{ik} = \sqrt{\varepsilon_{ii}\varepsilon_{kk}}$ and $R_{\min,ik} = \frac{1}{2}(R_{\min,ii} + R_{\min,kk})$.
>
> Given that the number of atom types is finite the authors precompute, for each grid point j, a vector $U_{LJ}^{j,\alpha}$ that stores the potential energy values for all atom types α. Ligand atoms are then associated with a one-hot vector w_i^α to select the corresponding van der Waals energy via a scalar product with the precomputed grid vector.

Collectively, these features define the weighted graphs used in the model. For the ligand:

$$G_{\text{mol}} = \{N_{\text{mol}}, E_{\text{mol}}, W_{\text{mol}}^N, W_{\text{mol}}^E\},$$

where W_{mol}^N includes charge w_i^q, van der Waals type vector w_i^α, and W_{mol}^E stores bond distances.

For the pocket:

$$G_{\text{grid}} = \{N_{\text{grid}}, E_{\text{grid}}, W_{\text{grid}}^N, W_{\text{grid}}^E\},$$

where node weights W_{grid}^N encode precomputed Coulomb w_j^q, van der Waals $w_{j,\alpha}^{\text{vdw}}$, , while edge weights W_{grid}^E represent spatial distances.

This enriched representation enables the formulation of an interaction-aware QUBO objective function that incorporates geometric compatibility along with biologically relevant interaction terms.

To incorporate physicochemical interactions into the optimization framework, additional terms are introduced into the geometric Hamiltonian of [1], among which is the interaction term investigated in this thesis. Each contribution is weighted and encoded using precomputed atom-level or grid-level descriptors.

The electrostatic interaction between a ligand atom i and a pocket point j is modeled as:

$$H_{\text{el}} = \sum_{i \in G_{\text{mol}}} \sum_{j \in G_{\text{grid}}} w_i^q \, w_j^{\text{el}} \, x_{ij}^2,$$

where w_i^q is the atomic charge of ligand atom i and w_j^{el} is the Coulomb potential at grid node j.

The van der Waals contribution is computed using the scalar product between the ligand's atom-type vector w_i^α and the Lennard-Jones potential vector at the grid node:

$$H_{\mathrm{vdw}} = \sum_{i \in G_{\mathrm{mol}}} \sum_{j \in G_{\mathrm{grid}}} w_i^\alpha \cdot w_j^{\mathrm{vdw}} \, x_{ij}^2 .$$

Combining all contributions, the full Hamiltonian becomes:

$$H = H_{\mathrm{geom}} + H_{\text{Ph-Ch}} = \lambda_1 H_{\mathrm{geom}} + \lambda_2 H_{\mathrm{el}} + \lambda_3 H_{\mathrm{vdw}} + \lambda_4 H_\pi , \tag{2.1}$$

where $\lambda_1, \ldots, \lambda_4$ are tunable parameters controlling the influence of each physicochemical term during the annealing process and H_π is the interaction term that models $\pi-$stacking interactions, which represents the core focus of this work.

2.2 Theoretical Nature of the π-Stacking Interaction

π-Stacking interactions are a class of noncovalent interactions occurring between aromatic systems, such as benzene rings, polycyclic aromatic hydrocarbons, and aromatic residues in biomolecules. The term π-stacking derives its name from the involvement of π electrons - those occupying delocalized molecular orbitals perpendicular to the plane of the aromatic ring. When two such systems are brought into proximity, favorable interactions between these π electron clouds can occur, leading to stacked arrangements that are energetically stabilized. These interactions play a fundamental role in a wide range of phenomena, including molecular recognition, protein folding, DNA stability, and the self-assembly of supramolecular structures. Although initially explained through purely electrostatic models, our understanding of π-stacking has evolved considerably, owing to the development of more refined quantum chemical approaches and a better grasp of the interplay between attractive and repulsive forces acting at the molecular level.

2.2.1 Phenomenological Description and Energy Components

From a geometrical standpoint, π-stacking interactions commonly occur in three principal configurations, particularly well exemplified by the benzene dimer, as shown in Figure 2.1. These can be examined along a parallel-displacement or “sliding” coordinate, either in a parallel or a perpendicular orientation between the monomers.

- *Parallel-displaced* (or slip-stacked), where the two aromatic rings are oriented in a roughly parallel fashion, but laterally offset. This geometry allows for significant surface contact while avoiding direct superposition of the electron clouds. The typical interplanar distance in this arrangement is approximately 3.4/3.5 Å while the displacement is

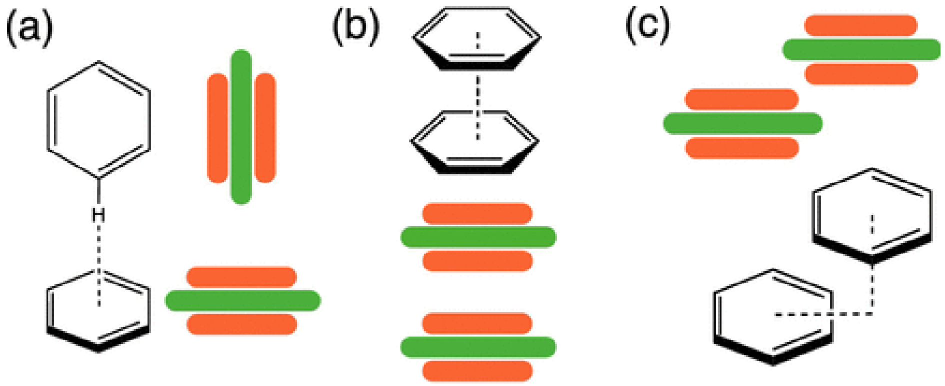


Figure 2.1: The three most common geometries of the benzene dimer: (a) T-shaped, (b) sandwich (cofacial) configuration , and (c) parallel-displaced. Image adapted from [23].

usually 1.7 Å, corresponding to a local energy minimum.

- *T-shaped*, where one ring lies approximately perpendicular to the plane of the other, forming a "T" configuration. The hydrogen atoms of one ring often point toward the center of the π-cloud of the other, leading to a favorable spatial arrangement. In this case, the center of one monomer is typically located about 5.0 Å from the plane containing the other ring.
- *Sandwich* (cofacial), where the two rings are aligned face-to-face with minimal offset. Despite maximizing the overlap between π-systems, this configuration is relatively rare due to strong short-range repulsion. It corresponds to a saddle point with an interplanar distance of about 3.8 Å.

Each of these configurations is stabilized or destabilized by a combination of physical forces. Although these forces will be discussed in greater detail in later sections, we introduce them briefly here to define their fundamental nature:

- *Electrostatic interactions*, arising from the distribution of static or induced charges within molecules. Aromatic rings, such as benzene, possess quadrupole moments resulting from the anisotropic distribution of π-electrons.

- *London dispersion forces*, a form of van der Waals interaction, result from transient fluctuations in electron density that create instantaneous dipoles.
- *Pauli (exchange) repulsion*, a short-range quantum mechanical repulsion that occurs when the electron clouds of closed-shell systems overlap, due to the Pauli exclusion principle.
- *Induction effects*, involving the polarization of the electron cloud of one molecule in response to the electrostatic field of another, generating a stabilizing interaction. This contribution becomes more significant in polar or heteroaromatic systems.

These energy terms do not act independently but in concert, generating a complex potential energy surface that governs the preferred geometry and stability of the interacting dimers.

2.2.2 Electrostatic Model and the Hunter–Sanders Framework

One of the most influential early models for describing π-stacking interactions is the electrostatic framework introduced by Hunter and Sanders [24]. This model attributes the preferred geometries of aromatic dimers to quadrupole–quadrupole electrostatics. In particular, it views benzene and similar aromatic systems as possessing a positively charged σ-framework sandwiched between two negatively charged π-clouds. Within this picture, electrostatic repulsion between the delocalized π-electrons of two cofacial rings renders the face-to-face arrangement unfavorable. Instead, offset geometries such as the parallel-displaced and T-shaped conformations emerge as means to minimize repulsive interactions.

The Hunter–Sanders model was introduced as an improvement over earlier, less satisfactory explanations. The *solvophobic model* [25], for instance, suggested that stacking is driven by entropic effects in solution; however, this theory fails to account for experimental observations in nonpolar solvents and often predicts maximal π–π overlap, which is not consistent with observed geometries. Another hypothesis, the *electron donor–acceptor (EDA) model* [26], links stacking to charge-transfer interactions. Yet this mechanism is only applicable to specific heterodimers and does not generalize to neutral aromatic systems like the benzene dimer. Similarly, the *atomic charge model* [27], based on partial atomic charges, fails to reproduce correct interaction energies and often underestimates binding by an order of magnitude.

By contrast, the Hunter–Sanders electrostatic model provides an intuitive and computationally simple way to rationalize geometrical preferences. In their formulation, the total interaction energy is decomposed into an electrostatic term and an attractive van der Waals (vdW) component, with induction considered negligible. The vdW term is not geometry-selective and does not explain why offset geometries are favored. Electrostatics, in this model, dominate the directional behavior of the interaction.

Subsequent studies, such as that by Swart et al. [28], have supported the importance of electrostatic interactions in π-stacking. Using a detailed energy decomposition analysis, these authors confirmed that quadrupolar electrostatics play a critical role in stabilizing the observed geometries, alongside Pauli repulsion and bonding orbital interactions. The Hunter–Sanders model remains a cornerstone of supramolecular chemistry, frequently cited for its conceptual clarity and predictive value.

2.2.3 Recent Developments and the Limitations of the Electrostatic Paradigm

Despite its success, the Hunter–Sanders model has faced increasing scrutiny from high-level quantum chemical studies. In particular, recent work by Carter-Fenk and Herbert [29] challenges the notion that electrostatics alone dictate the slip-stacked arrangement of aromatic dimers. Their findings reveal that classical quadrupole-based models may misrepresent the true physical driving forces behind π-stacking, especially at the short intermolecular distances typical of these interactions.

Using *ab initio* methods that include charge penetration effects and accurate electron correlation, the authors show that electrostatics can in fact be attractive in cofacial geometries—contradicting the predictions of the Hunter–Sanders model. Moreover, studies on substituted dimers, such as C_6H_6/C_6F_6, demonstrate that both electron-withdrawing and electron - donating groups tend to increase interaction strength, regardless of their effect on quadrupole moments. This observation is incompatible with a model based solely on electrostatic potential maps.

These inconsistencies have led to a revised interpretation: the primary forces controlling π–π geometries are not electrostatics and dispersion, but rather the competition between London dispersion and Pauli repulsion. In this view, offset geometries such as the slip-stacked configuration emerge not to minimize quadrupolar repulsion, but to balance the attractive dispersion with the repulsive exchange forces that dominate at close contact. A simple van der Waals potential model, grounded in this idea, aligns closely with computational data and offers a more universal explanation for aromatic stacking behavior.

This vdW-based perspective explains why parallel-displaced configurations are consistently favored across systems with vastly different electrostatic profiles. It also accounts for the observation that electrostatics play only a minor role at close range, where charge densities overlap significantly. Consequently, the modern understanding of π-stacking places greater emphasis on quantum mechanical effects like electron correlation and exchange repulsion, challenging classical electrostatic frameworks.

2.3 Computational Chemistry: An Overview

Computational chemistry is a discipline that employs theoretical models and numerical algorithms to address chemical problems through computer-based simulations. It enables the prediction of molecular properties, electronic structures, chemical reactivity, and intermolecular interactions, thereby reducing the reliance on laboratory experiments, which are often costly or difficult to perform. By integrating concepts from quantum mechanics, molecular mechanics, and statistical dynamics, computational chemistry represents a multidisciplinary field at the intersection of chemistry, physics, and computer science.

In recent decades, advances in theoretical methods and the continuous growth in computational power have significantly expanded the scope of computational chemistry. It is now possible to simulate complex systems, including crystalline materials, biomolecules, catalysts, and molecular systems in both gas and condensed phases. Computational results often show good agreement with experimental data, making this discipline an essential tool in both academic research and industrial innovation.

To investigate the electronic structure and properties of molecular systems, a wide range of computational methods has been developed. Among these, two main families dominate the landscape of quantum chemistry: *ab initio* methods [30] and Density Functional Theory (DFT) [31]. These approaches differ in their theoretical foundations, computational cost, and range of applicability, yet both play a central role in modern computational investigations.

2.3.1 Density Functional Theory

Density Functional Theory (DFT) [31] is one of the most widely used theoretical framework in computational quantum chemistry and condensed matter physics for studying the electronic structure of many-body systems. Its central concept is to describe the properties of an interacting electron system in terms of its total electron density $\rho(\mathbf{r})$, rather than the more complex many-electron wavefunction. This reformulation dramatically reduces computational complexity, as the electron density depends only on three spatial coordinates, regardless of the number of electrons in the system.

The theoretical foundation of DFT is provided by the two Hohenberg-Kohn theorems [32], formulated in 1964. The first theorem establishes that the ground-state properties of a many-electron system are uniquely determined by its ground-state electron density, implying that all observables, including the total energy, can be expressed as functionals of the density. The second theorem introduces a variational principle, stating that the correct ground-state electron density minimizes the total energy functional.

However, the exact form of the functional that maps the electron density to the total energy remains unknown, particularly for the exchange-correlation component, which encapsulates all many-body quantum effects beyond classical electrostatics and the non-interacting kinetic energy.

To make DFT practically solvable, the Kohn-Sham [33] formalism introduces a fictitious system of non-interacting electrons that reproduces the exact ground-state density of the original interacting system. This leads to the Kohn-Sham equations, a set of single-particle Schrödinger-like equations that are solved iteratively within a self-consistent field (SCF) procedure.

The Kohn-Sham Equations

The Kohn-Sham (KS) equations are a cornerstone of modern computational chemistry, providing a practical and efficient way to apply DFT to many-electron systems. Introduced by W. Kohn and L. J. Sham in 1965 [33], this approach replaces the complex, interacting N-electron problem with an equivalent system of N non-interacting electrons moving in an effective potential.

In this formulation, the total energy of the system is expressed as a functional of the electron density $\rho(\mathbf{r})$, incorporating contributions from the kinetic energy of the non-interacting system, the classical Coulomb (Hartree) interaction, the external potential (typically due to nuclei), and the exchange-correlation energy:

$$E[\rho] = T_s[\rho] + E_{\text{ext}}[\rho] + E_{\text{H}}[\rho] + E_{\text{xc}}[\rho], \tag{2.2}$$

where $T_s[\rho]$ is the kinetic energy of the non-interacting system, $E_{\text{ext}}[\rho]$ is the interaction with the external potential, $E_{\text{H}}[\rho]$ is the classical electron-electron repulsion, and $E_{\text{xc}}[\rho]$ accounts for exchange and correlation effects.

The Kohn-Sham equations are derived by minimizing this total energy functional under the constraint that the electron density integrates to the correct number of electrons. This leads to the following set of equations:

$$\left[-\frac{1}{2}\nabla^2 + V_{\text{eff}}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}), \tag{2.3}$$

where $\psi_i(\mathbf{r})$ are the Kohn-Sham orbitals, ϵ_i are the corresponding orbital energies, and $V_{\text{eff}}(\mathbf{r})$ is the effective potential, defined as:

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ext}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{xc}}(\mathbf{r}), \tag{2.4}$$

with $V_{\text{ext}}(\mathbf{r})$ representing the external potential, $V_{\text{H}}(\mathbf{r})$ the Hartree potential, and $V_{\text{xc}}(\mathbf{r})$ the exchange-correlation potential, obtained as the functional derivative of the exchange-correlation energy:

$$V_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\rho]}{\delta\rho(\mathbf{r})}. \tag{2.5}$$

Despite the approximations involved in modeling the exchange-correlation energy, the Kohn-Sham equations, when combined with suitable functionals, have proven highly effective in accurately describing the ground-state properties of atoms, molecules, and solids at a reasonable computational cost.

2.3.2 *Ab Initio* Methods

Ab initio methods [30] constitute a class of quantum chemical approaches that aim to solve the electronic Schrödinger equation starting solely from first principles, without the use of empirical parameters. These methods provide systematically improvable approximations to the electronic structure of molecular systems and are particularly valued for their ability to offer detailed insights into the nature of electronic interactions.

In the context of this work, special attention is given to two distinct yet complementary *ab initio* approaches: the Hartree-Fock (HF) method [34] [35], introduced in Section 2.3.2, and Symmetry-Adapted Perturbation Theory (SAPT) [36], introduced in Section 2.3.2, including its extended variant XSAPT. The Hartree-Fock method serves as the foundation for many post-Hartree-Fock techniques and provides reference wavefunctions for perturbative treatments. SAPT-based methods, by contrast, are specifically designed to compute intermolecular interaction energies and offer a physically meaningful decomposition of these interactions into distinct components such as electrostatics, exchange, induction, and dispersion.

The following sections present the theoretical foundations and practical aspects of both Hartree-Fock and SAPT-based approaches, emphasizing their respective roles in the accurate modeling of non-covalent interactions.

Hartree-Fock

The Hartree-Fock (HF) [34] [35] method is one of the most fundamental and widely used *ab initio* approaches in quantum chemistry for determining the electronic structure of atoms, molecules, and solids. Developed by D. R. Hartree and later refined by V. Fock, the method provides an approximate solution to the many-electron Schrödinger equation using a mean-field approach.

In Hartree–Fock theory, the total wavefunction of a molecular system is approximated by a single Slater determinant [37], representing the state of a non-interacting electronic system. This approximation is obtained by applying the variational principle to minimize the system's total energy. This procedure minimizes the total electronic energy with respect to the orbitals, subject to the constraint that they remain orthonormal. The resulting set of eigenvalue equations, known as the Hartree-Fock equations, describes how each electron moves in the average field generated by all other electrons:

$$H_{\mathrm{HF}}\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}), \tag{2.6}$$

where H_{HF} is the Hartree-Fock Hamiltonian, $\psi_i(\mathbf{r})$ is the i-th molecular orbital, and ϵ_i is its associated energy.

The Hartree-Fock Hamiltonian includes both the one-electron Hamiltonian - accounting for kinetic energy and electron-nucleus attraction - and a mean-field representation of the electron-electron repulsion, captured by the Coulomb and exchange terms. Notably, the Hartree-Fock method does not include dynamic electron correlation effects, making it a mean-field approximation.

The solution of the Hartree-Fock equations is obtained through a self-consistent field (SCF) procedure, where an initial guess for the molecular orbitals is iteratively refined until convergence is achieved.

Despite its limitations, especially regarding the neglect of electron correlation, the Hartree-Fock method remains a cornerstone of electronic structure theory. It serves as the reference point for more advanced post-Hartree-Fock methods such as Møller-Plesset perturbation theory (MP2), Configuration Interaction (CI), Coupled Cluster (CC), and perturbative approaches like SAPT, which systematically improve upon the Hartree-Fock approximation by incorporating electron correlation effects.

SAPT and XSAPT

Symmetry-Adapted Perturbation Theory (SAPT) [36] is a well-established *ab initio* method specifically developed for the study of intermolecular interactions. Unlike traditional supermolecular approaches, which compute interaction energies as simple differences between total energies of monomers and complexes, SAPT provides a physically meaningful decomposition of the interaction energy into distinct components. This decomposition offers valuable insights into the nature of non-covalent interactions such as π-stacking, hydrogen bonding, and van der Waals forces.

The total SAPT interaction energy up to second order can be expressed as:

$$E_{\text{int}} = E^{(1)}_{\text{elst}} + E^{(1)}_{\text{exch}} + E^{(2)}_{\text{ind}} + E^{(2)}_{\text{exch-ind}} + E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exch-disp}}, \tag{2.7}$$

where the terms correspond to:

- $E^{(1)}_{\text{elst}}$: First-order electrostatics, arising from interactions between charge distributions of the isolated fragments;
- $E^{(1)}_{\text{exch}}$: First-order exchange, representing Pauli repulsion due to antisymmetrization of the wavefunction;
- $E^{(2)}_{\text{ind}}$: Second-order induction, accounting for the polarization of one fragment induced by the other;
- $E^{(2)}_{\text{exch-ind}}$: Exchange correction to the induction energy;
- $E^{(2)}_{\text{disp}}$: Second-order dispersion, resulting from correlated electronic fluctuations between fragments;
- $E^{(2)}_{\text{exch-disp}}$: Exchange correction to the dispersion energy.

To improve both the accuracy and scalability of SAPT for larger systems, the Extended SAPT (XSAPT) method has been developed. XSAPT introduces several key modifications while retaining the physically transparent decomposition of the interaction energy. One major improvement lies in the treatment of induction: XSAPT employs an electrostatic embedding scheme, allowing each fragment to polarize in the presence of the static electric field generated by the surrounding fragments. This more realistic description of the molecular environment enhances the accuracy of induction energy calculations, particularly in large and complex systems.

Another important feature of XSAPT is the optional incorporation of many-body dispersion effects via the MBD (Many-Body Dispersion) model. In the XSAPT+MBD framework, the standard SAPT dispersion term is replaced by the MBD contribution:

$$E_{\text{disp}} = E_{\text{MBD}}, \tag{2.8}$$

providing a more accurate description of long-range non-local dispersion interactions.

The calculation of the XSAPT interaction energy typically requires at least two distinct computational steps. The first is a conventional second-order SAPT calculation on the entire system without charge embedding, yielding the electrostatic, exchange, and (if MBD is not used) dispersion terms. The second step involves performing SAPT calculations with charge embedding for each monomer pair, allowing the induction energy to be corrected for the influence of the surrounding environment.

The final induction energy in XSAPT is obtained by summing the original SAPT induction term and the difference between the total SAPT energies with and without embedding:

$$E_{\text{ind}} = E_{\text{ind}}^{(2)} + E_{\text{exch-ind}}^{(2)} + \delta E_{\text{HF}}, \tag{2.9}$$

where δE_{HF}, known as the δ_{HF} correction, is evaluated through a Hartree-Fock calculation on the dimer. This term estimates induction effects beyond second order in perturbation theory and is particularly important when strong polarization effects are present.

The overall XSAPT energy, especially when combined with the MBD correction, provides a computationally efficient yet physically detailed description of intermolecular interactions. By allowing for environmental polarization and many-body dispersion, XSAPT extends the applicability of SAPT to larger molecular systems, such as molecular clusters and biomolecular complexes, while preserving the desirable feature of energy decomposition into physically interpretable components. For a more comprehensive overview of the theoretical background underlying the SAPT method, the reader is referred to reference [36].

Chapter 3

High-Performance Computing Simulations for QUBO Model Development

To quantify the intensity of the π-stacking interaction, we conducted an extensive set of *ab initio* computational chemistry simulations, leveraging the resources of the *Leonardo* supercomputer (1.2). We chose to pursue this research direction because, to the best of our knowledge, a systematic analysis of π-stacking energy as a function of the relative positions of the two molecules had not yet been performed. In particular, no explicit formula exists in the literature that allows one to determine the interaction energy directly from the geometric configuration of the involved molecules.

For this reason, it was necessary to rely on numerical simulations in computational chemistry in order to obtain a quantitative characterization of the phenomenon. These simulations allowed us to analyse in detail how variations in position and orientation of the molecules influence the interaction energy, thus providing a solid empirical basis for defining the energetic contribution associated with the π-stacking interaction.

Our analysis initially followed the work of Carter et al. [29]. Subsequently, we referred to a more recent publication by the same authors [23], which offered a robust starting point for the simulations carried out in this study.

These works are considered authoritative references in the study of π-stacking interactions between aromatic rings and were thus deemed an ideal starting point for validating our computational approach. Our initial objective was to accurately reproduce the interaction energies reported in the article using the same molecular setup. Specifically, our aim was to replicate the results related to the interaction energy between two benzene molecules in the parallel-displaced configuration, with an interplanar vertical distance of 3.4 Å and a variable displacement in the interval $[-2.9, 2.9]$. Obtaining results consistent with those published allowed us to validate the setup of our simulations from both a theoretical and numerical perspective, ensuring that the model was capable of correctly capturing the physical characteristics of the phenomenon.

Only after completing this validation phase, we extended the study to geometric configurations different from those analysed in [23], exploring new orientations and distances between the molecules with the goal of characterizing the interaction energy more comprehensively as a function of their relative spatial arrangement.

For our initial simulations, we first chose to use the software package Quantum ESPRESSO [38], which is one of the most widespread an reliable code for DFT, as it allows for results to be obtained relatively quickly and with limited use of computational resources. We considered the use of Quantum ESPRESSO appropriate in the preliminary phase, since density functional theory based methods generally provide a good approximation of the results achieved with more accurate, but also more expensive, *ab initio* techniques.

However, the results obtained using Quantum ESPRESSO turned out to be inconsistent with both those reported in [23] and the values theoretically expected based on the literature. This highlighted the need to adopt a more sophisticated approach, capable of treating non-covalent interactions with greater accuracy.

3.1 QE Simulations

Quantum ESPRESSO is an integrated suite of open-source codes for modeling materials from the perspective of density functional theory (DFT), introduced in Section 2.3.1. It is widely used to study crystalline materials, surfaces, structural defects, phonon vibrations, and many other physical properties.

An input file for Quantum ESPRESSO consists of a series of structured sections, each defining physical parameters, computational settings, and data related to the system to be simulated. The syntax is simple and block-based, with each block preceded by the symbol & (e.g., `&CONTROL`, `&SYSTEM`, ...) and terminated by a slash `/`.

3.1.1 Code Settings

For each simulations, we selected the following parameters:

- `calculation = 'scf'`, as we intended to perform a self-consistent field (SCF) calculation, which determines the electronic ground-state density that satisfies the Kohn–Sham equations:

  $$\hat{F}[\rho]\,\varphi_i = \varepsilon_i\,\varphi_i, \tag{3.1}$$

  where $\hat{F}[\rho]$ is the Kohn–Sham operator (dependent on the electronic density ρ), $\{\varphi_i\}$ are the Kohn–Sham orbitals, and ε_i the corresponding eigenvalues. The electronic density ρ is iteratively updated until self-consistency is reached, i.e., until the input and output densities match within a predefined tolerance threshold.

- `Pseudopotentials`: we used norm-conserving pseudopotentials because, in quantum mechanical calculations, electronic wavefunctions exhibit strong oscillations near the nucleus, where core electrons reside. Accurately describing these oscillations requires very large and dense basis sets, significantly increasing computational cost. To overcome this difficulty, *pseudopotentials* are used to replace the full Coulomb potential with a smoother effective potential that only acts on valence electrons.

Norm-conserving pseudopotentials are a specific class designed to preserve the *norm* of the wavefunction near the nucleus. This means the electronic charge density in that region is the same for both the all-electron and the pseudized wavefunction. This condition ensures accurate reproduction of physical properties such as energy, forces, and charge distribution.

The main advantage of norm-conserving potentials lies in their physical reliability and transferability: they can be used across different chemical environments without modification. However, they typically require relatively high kinetic energy cutoffs (`ecutwfc`) to achieve numerical convergence, making the calculations still computationally demanding.

- `ibrav = 0` and `CELL_PARAMETERS=[15.0 0.0 ...]` were used because the system consists of two isolated benzene molecules without any periodic crystalline structure. Quantum ESPRESSO inherently performs calculations with periodic boundary conditions (PBC) in all directions. Even in simulations of isolated molecules, a simulation cell must be defined, and it is assumed to repeat periodically.

  By setting `ibrav = 0`, we discard the use of predefined cell types (cubic, triclinic, hexagonal, etc.) and manually specify the three cell vectors in the `CELL_PARAMETERS` section. This allows us to construct an arbitrary simulation cell, large enough to minimize spurious interactions between periodic images of the molecules.

  This choice is essential for non-periodic systems to avoid artifacts due to image-image interactions. A cell that is too small would cause overlap between replicated images and introduce significant errors in energy and electron density calculations.

- `vdw_corr` is used to account for van der Waals interactions. We tested

both `‘grimme-d3’` and `‘MBD’` corrections. `‘grimme-d3’` applies an empirical damping correction with negligible computational cost and pre-calibrated parameters.
`‘MBD’` uses a many-body model of coupled harmonic oscillators with density-based electronic screening, including higher-order contributions. Grimme-D3 is ideal for fast simulations on isolated molecules, while MBD offers greater accuracy for extended systems at higher computational cost. We used both schemes, and the results obtained are reported in the next section.

- `ecutwfc=110 Ry` sets the kinetic energy cutoff for the plane-wave basis of the electronic wavefunctions, ensuring convergence. `ecutrho=880` sets the cutoff for the charge density and exchange-correlation potentials. Typically, ecutrho is set at $\approx 8 \times$ ecutwfc to capture high-frequency components. These values were chosen empirically through total energy convergence tests to balance numerical accuracy and computational cost.

- `conv_thr = 1.0d-8` sets the SCF convergence threshold to 10^{-8} Ry, ensuring sufficiently high energy accuracy.

- `K_POINTS`: Sets the grid for the calculation. For isolated molecules, a single-point grid `1 1 1 0 0 0` is used.

3.1.2 Comparison Between Obtained and Expected Results

Below, we report the results obtained from simulations performed using the *Quantum ESPRESSO* software, concerning the *parallel-displaced* configuration between two benzene molecules. Specifically, the vertical distance between the molecular planes is fixed at 3.4 Å, while the horizontal displacement varies within the interval $[-2.9, 2.9]$ Å.

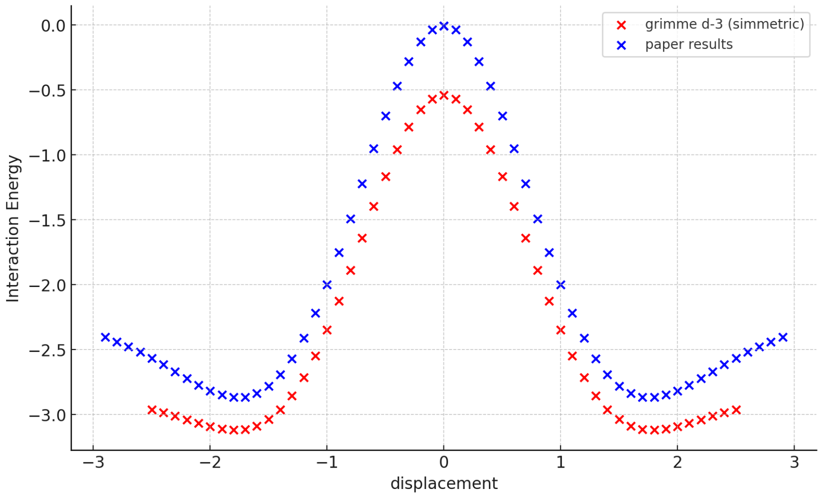


Figure 3.1: Comparison between results from Quantum ESPRESSO simulations (red) and data from [23] (blue), related to π-stacking interaction in the *parallel-displaced* configuration.

In Figure 3.1, we compare the interaction energy values reported in [23] (in blue) with those obtained from our simulations (in red).

As observed, the curve obtained from our simulations is slightly shifted downward compared to the one reported in the literature. Moreover, the positions of minima and maxima do not align perfectly, and the tails of our curve show a gentler slope than those of the reference curve.

In Figure 3.2, we instead report the interaction energy values obtained using the `vdw_corr = 'MBD'` parameter within the interval $[-1.6, 1.6]$ Å. We observe that with this correction, the maximum of the curve coincides with the reference, but the energy values at the minima are not consistent with theoretical expectations or with those reported by [23].

As evidenced by the results shown above, the energy values obtained from the *Quantum ESPRESSO* simulations do not reproduce the data reported in [23] with sufficient accuracy. In particular, significant discrepancies are observed in the position and magnitude of the interaction energy minima and maxima, which play a central role in our analysis, as they define the most stable and relevant spatial configurations for modeling purposes.

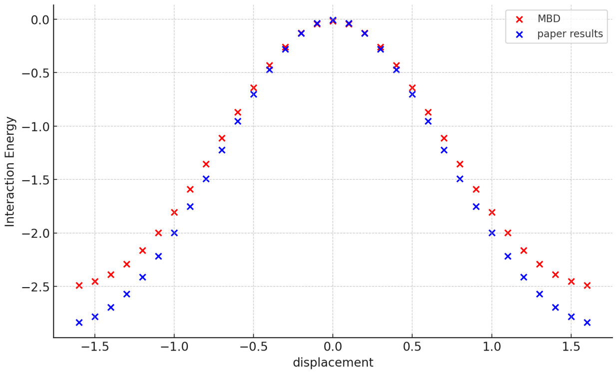


Figure 3.2: Comparison between results from MBD-corrected simulations (red) and data from [23] (blue).

For this reason, we decided to perform a second study of the interaction energy using the *Q-Chem* software, with the aim of obtaining a more reliable estimate of the energetic values and verifying the consistency of the results with those reported in the literature.

3.2 Q-Chem Simulations

Q-Chem is one of the most advanced commercial software packages for quantum chemistry, particularly suited for studying the electronic structure of molecules in the gas phase, liquid phase, or in solution. Compared to Quantum ESPRESSO, Q-Chem is more focused on isolated molecular systems and small organic molecules, although it is also capable of treating larger systems using approximate methods. Q-Chem includes a wide range of computational techniques, from DFT to post-Hartree-Fock methods (such as MP2, CCSD, EOM-CC), as well as advanced tools for studying excited states, electronic transitions, and non-covalent interactions.

Specifically, to compute the π-stacking interaction energy between two benzene rings, Ref.[23] uses a method known as XSAPT+MBD, introduced in Section 2.3.2.

3.2.1 Code Settings

The input file for running an XSAPT calculation is organized into three main sections: `$rem`, `$xpol`, and `$sapt`. The `$rem` section includes the general calculation parameters, such as job type, functional, basis set, and convergence criteria. The `$xpol` section specifies the treatment of fragment polarization, while `$sapt` defines the settings for energy decomposition according to SAPT theory. For jobs with and without embedding, an additional `$lrc_omega` section sets the value of ω used in the range-separated functional.

The key parameters used in the SAPT job with embedding are:

- `jobtype xsapt`: specifies that an XSAPT analysis is to be performed.
- `method lrc-wpbe`: sets the density functional to LRC-ωPBE, suitable for treating non-local interactions.
- `basis def2-tzvpd`: selects the `def2-TZVPD` basis set, which includes diffuse and polarization functions.
- `lrc_dft true`: enables the use of a range-separated functional in the DFT context.
- `scf_guess autosad`: uses the AutoSAD method for the initial SCF guess, which is robust for multi-fragment systems.

In the `$xpol` section, the parameter `embed charges` activates electrostatic embedding, meaning fragments are polarized by the field generated by the charges of other fragments. The option `dft-lrc` indicates that polarization is calculated using the range-separated functional specified in `$rem`.

In the `$sapt` section, the SAPT calculation is performed using the atomic orbital formalism (`algorithm ao`) and limited to second-order perturbation theory (`order 2`). A projected basis is used for decomposition (`basis projected`), and dispersion is treated via the `MBD` (Many-Body Dispersion) correction, which improves the description of non-local interactions.

Finally, the `$lrc_omega` section sets the value $\omega = 0.34$ bohr^{-1}, obtained using the *global density-dependent (GDD) tuning procedure*.

In the SAPT job without embedding, the main difference lies in the treatment of electrostatic embedding in the `$xpol` section, where `embed none` is specified. This implies that no electrostatic embedding is applied during the polarization calculation of the fragments, unlike the previous job where `embed charges` was used. As a result, the induction energy does not account for the field generated by charges of the other fragments. All other parameters—including the functional, basis set, MBD correction, and ω value—remain unchanged.

The following job is used to compute the δ_{HF} correction to the induction energy within an extended XSAPT procedure. This calculation aims to account for higher-order induction effects not captured by standard perturbative methods. It is therefore an optional but recommended step to improve the accuracy of the π-stacking interaction in critical configurations. For this job, the parameters in the `$rem` section are:

- `jobtype xsapt`: specifies that this is an XSAPT job, although in this case it is used solely to obtain the δ_{HF} correction;
- `method HF`: sets Hartree-Fock theory for computing the total electronic energy;
- `basis 6-311G`: specifies the basis set, which is simpler than those used previously, since the δ_{HF} calculation does not require an extended basis to yield qualitatively correct results;

In the `$xpol` section, electrostatic embedding is disabled (`embed none`), as the δ_{HF} correction is computed without accounting for polarization induced by external fragments.

The `$sapt` section activates the actual δ_{HF} computation. Key directives include:

- `cphf`: enables solving the Coupled Perturbed Hartree-Fock (CPHF) response equations;
- `dscf`: activates the explicit calculation of the difference between the full SCF energy of the dimer and the sum of the SCF energies of the monomers, which constitutes the δ_{HF} correction;
- `algorithm ao`, `order 2`, `basis dimer`: specify that SAPT is carried out in atomic orbital representation, up to second order, using a basis centered on the dimer;
- `Dispersion MBD`: although this term does not directly affect δ_{HF}, it is still treated using the Many-Body Dispersion model.

Overall, this job provides the δ_{HF} contribution, which is added to the induction energy obtained from the standard XSAPT jobs, thus improving the overall accuracy of the interaction energy.

Finally, we used the `xsapt_data_collection.py` script to extract the relevant interaction energies (see [39]).

3.3 Simulation Results

In this section, we present the results obtained from the simulations carried out using the *Q-Chem* software. The code used for the calculations was kindly provided by the authors of [23], which allowed us to exactly reproduce the energy results reported in the reference article.

It is important to note that the atomic coordinates of the benzene fragments used in the simulations, including carbon and hydrogen atoms, were provided directly by the authors of [23]. All details regarding the geometric optimization of the monomers are available in the supplementary material attached to the article. Starting from these optimized geometries, we introduced controlled modifications to the relative positions of the two molecules in order to explore a wide range of spatial configurations and to analyze in

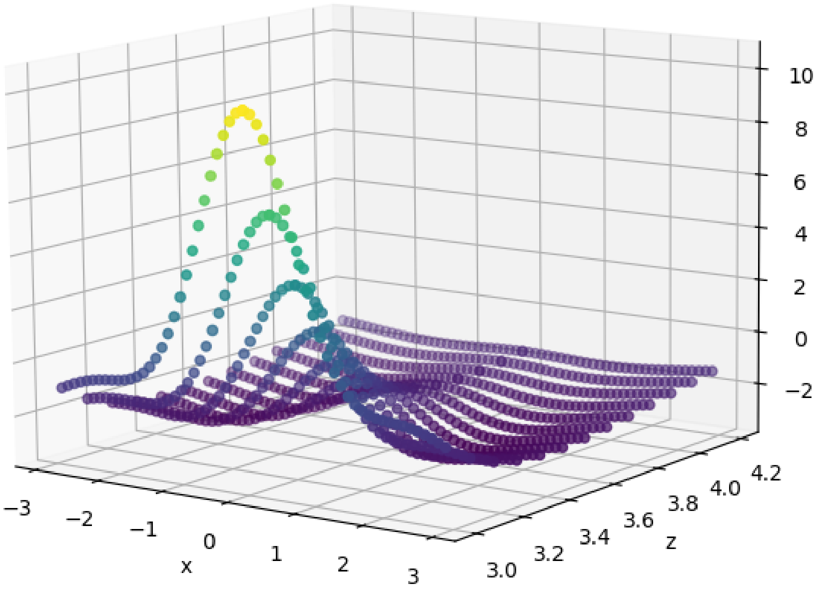


Figure 3.3: Interaction energy as a function of lateral displacement and vertical separation

greater detail the dependence of the π-stacking interaction on the distance and mutual orientation of the fragments.

3.3.1 parallel-displaced: translations

In this section, we present only the results obtained by translating the two molecules with optimized geometry. In particular, the translations were performed by shifting the center of one benzene fragment—and consequently all its atoms—via a rigid displacement, while keeping the other fixed. As a first step, we report the interaction energy values as a function of the vertical distance between the planes containing the two benzene molecules and the horizontal *displacement* between them. In Figure 3.3, x denotes the displacement (i.e., the horizontal distance between the centers of the two parallel molecules), and z is the vertical distance between the planes of the aromatic rings.

The first observation that emerges from the analysis of the plot is the perfect symmetry with respect to zero for the horizontal *displacement*. This behavior was expected, as it reflects the intrinsic symmetry of the system composed of two identical benzene molecules.

Moreover, it can be seen that for vertical distance values z less than 3.4 Å, the interaction energy increases very rapidly for small *displacement* values, indicating a strong sensitivity of the interaction to lateral displacement in the short-range regime. It is also observed that, regardless of the vertical distance z, the region of space corresponding to *displacement* values around 1.7 Å represents a minimum zone for the interaction energy. This indicates that such a lateral configuration is energetically favorable across the entire range of distances considered. Additionally, for large values of vertical distance z, the energy surface progressively flattens out, with the interaction energy gradually approaching zero. This behavior reflects the progressive cancellation of the π-stacking interaction, which becomes negligible as the distance between the molecular planes increases. In such configurations, the overlap of electronic densities and non-covalent interactions are significantly weakened, eventually disappearing beyond a certain threshold.

To ensure better readability of the plot, in Figure 3.4 we have removed from the visualization the energy values for $z = 3$ Å and $z = 3.1$ Å, which exhibit very high "peaks" that flatten the scale of the entire graph.

We then extended the simulations by also including translations of the molecule along the axis orthogonal to the displacement axis, which for consistency we denote by the letter y. This direction, distinct from the horizontal displacement, allows us to explore additional configurations. In particular, we chose to focus the new simulations on configurations characterized by displacement values between 1.0 and 2.9 Å and vertical distances z between 3.2 and 4.0 Å. Configurations outside these intervals were excluded from the analysis, as they are associated with interaction energies too weak to support a stable π-stacking interaction. We chose to perform the new simulations by considering y values between 0 and 1.6 Å. This interval was chosen empirically, in line with what was previously noted.

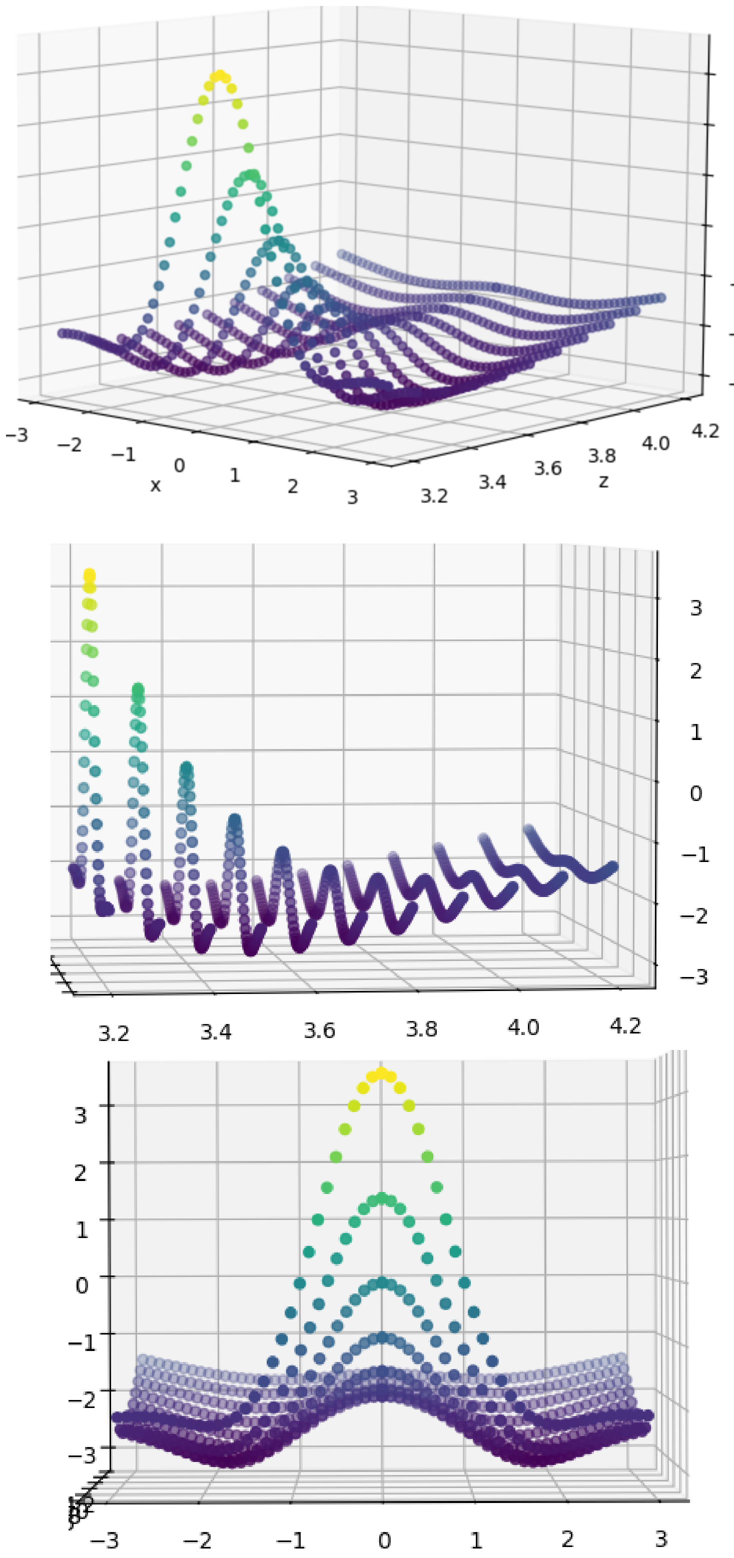


Figure 3.4: Interaction energy as a function of lateral displacement with vertical separation ≥ 3

We observed a strong correlation between energy, vertical distance, and planar radial distance, defined as

$$\sqrt{x^2 + y^2}, \tag{3.2}$$

i.e., the distance between the centers of the two molecules in the plane parallel to their aromatic rings.

In Figure 3.5 we show the plot representing the interaction energy as a function of the planar radial distance and the vertical distance between the molecular planes. In particular, this graph displays all values collected from the simulations. From the plot, it can be observed that in cases where the planar radial distance does not depend solely on the displacement along the x-axis, but also involves translations along the y-axis (i.e., for x values greater than 1), the points appear slightly scattered compared to the main curve. Nevertheless, the overall trend remains consistent, confirming that the dependence of the interaction energy on the planar radial distance retains a coherent structure.

To better display the results obtained, in Figure 3.6 (a) we consider only values of $z > 3.2$, thus excluding excessively high energy values that limit the clarity of the graphical representation. In 3.6 (b) and (c), we report the interaction energy as a function of the planar radial distance for a fixed vertical distance of $z = 3.5$ Å. Figure 3.6 (b) shows the full set of simulated values, while 3.6 (c) represents only configurations with energy $\leq$ -2 kcal/mol.

As can be seen, when we closely examine the plot, we observe more scattered values at the tails, while near the minimum the values are more concentrated. A fundamental aspect of our study concerns the definition of an energy threshold below which the π-stacking interaction can be considered sufficiently effective to be relevant for modeling purposes. To this end, we chose to adopt a threshold value of -2 kcal/mol. This choice is motivated by the fact that, in the literature, the typical interaction energy at equilibrium—such as in the *parallel displaced* arrangement of two benzene molecules—is around -3 kcal/mol. Our simulations show that this value is not fixed, but varies continuously with the relative position and orientation

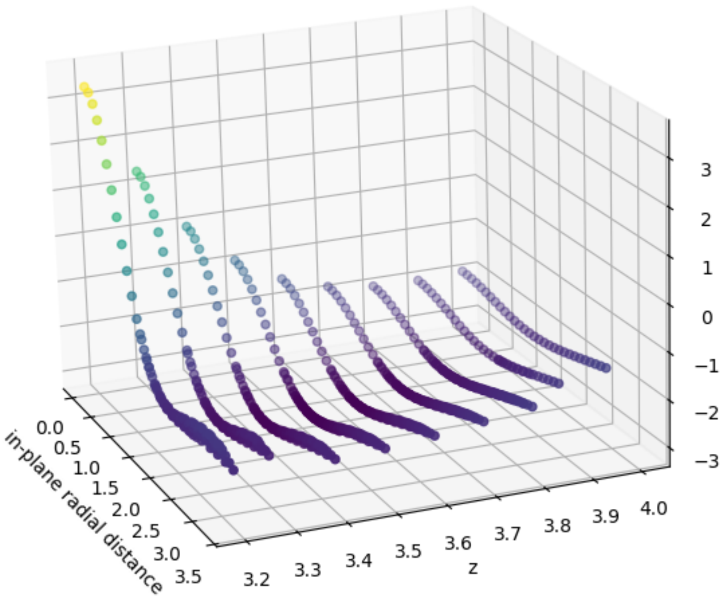


Figure 3.5: Interaction energy as a function of the planar radial distance and the vertical distance

of the aromatic fragments. In other words, the interaction is not present in a discrete manner but gradually weakens as a function of spatial translations and rotations. For this reason, it is necessary to introduce a conventional threshold that can consistently and rigorously distinguish spatial configurations that should be considered as effectively interacting from those where the interaction is negligible. Choosing -2 kcal/mol as the lower limit allows us to include in the model all configurations with significant interaction, while excluding those that, although showing a weak residual attraction, do not contribute substantially to the stabilization of the system.

In Figure 3.7 we show the plots of interaction energy as a function of planar radial distance and z, considering only the points with energy less than or equal to -2 kcal/mol.

As can be seen, by restricting the analysis to a limited energy interval, data dispersion becomes more evident and the overall trend deviates more from a smooth and well-defined curve. In particular, small fluctuations around the expected energy profile emerge more clearly, which were

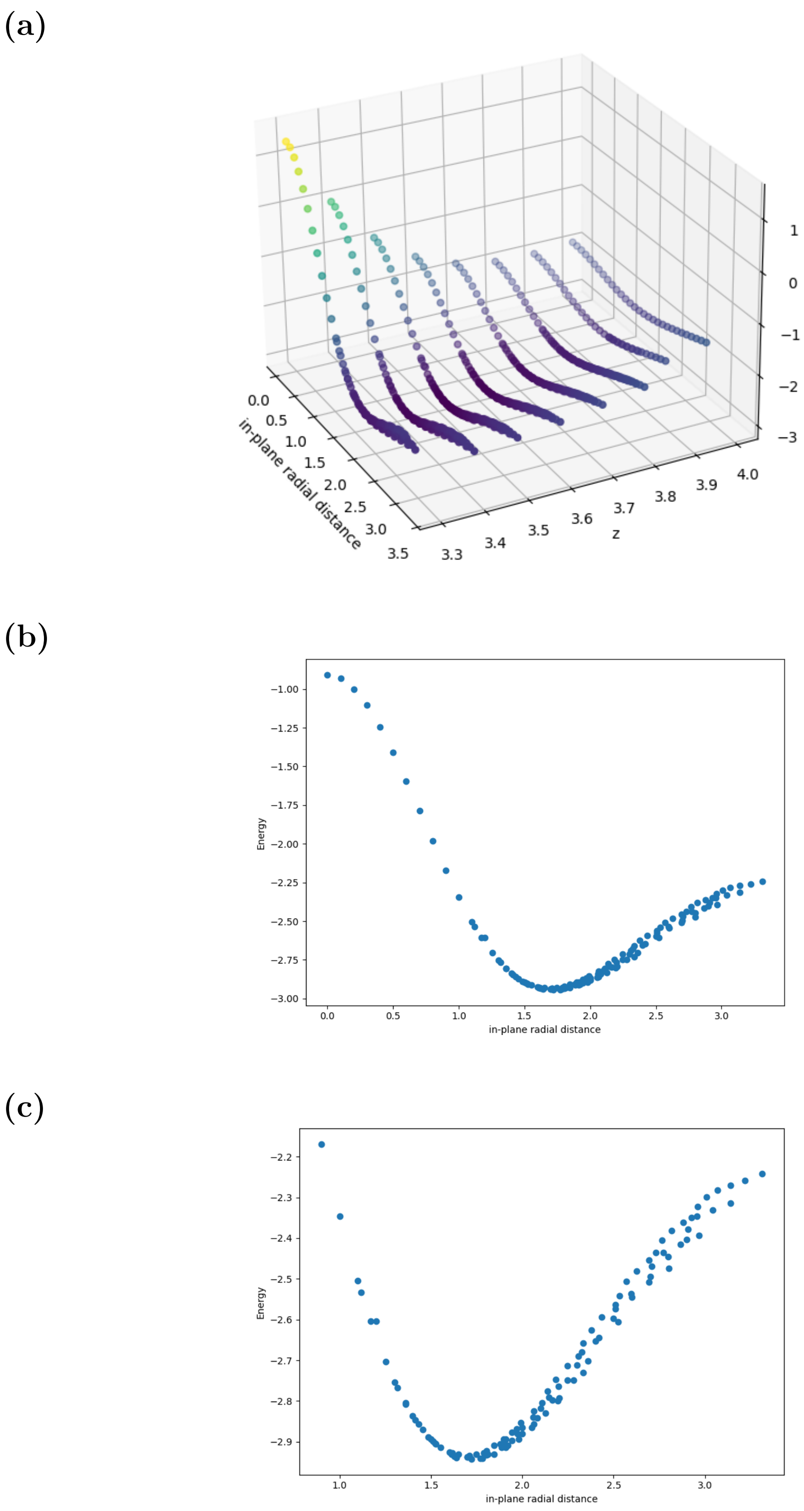


Figure 3.6: Interaction energy as a function of the planar radial distance with: (a) vertical distance > 3.2 Å; (b) fixed vertical distance $z = 3.5$ Å; (c) $z = 3.5$ Å and energy ≤ -2.

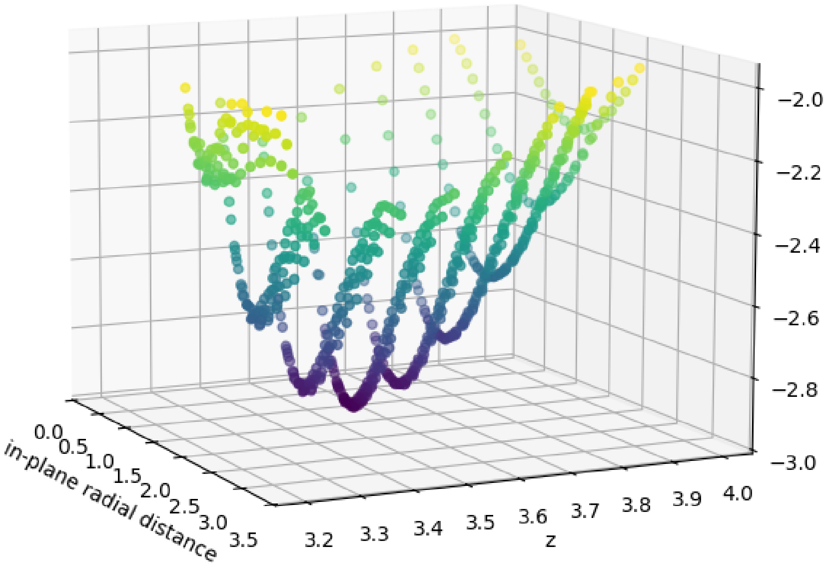


Figure 3.7: Interaction energy as a function of planar radial distance and z, with energy ≤ -2 kcal/mol

not immediately visible in the broader energy range.

This effect is especially pronounced for smaller vertical distance values z, where the interaction between the fragments is more intense and sensitive even to slight spatial variations. Moreover, the most significant deviations are observed for displacement values greater than the minimum of each curve, i.e., in regions where the interaction starts to weaken but is not yet entirely negligible. This behavior suggests that, outside the energy minimum, the dependence of the interaction energy on small changes in relative geometry becomes more complex and less predictable.

In Figure 3.8, we report the plots of interaction energy considering only configurations with vertical distance $z \geq 3.4$ Å. As can be observed, the energy surface in this case appears much more regular and well-defined.

For completeness, in Figure 3.9 we show the results obtained for the values $z = 3.2$, 3.3, and 3.4 Å.

In these cases, we observe a much more pronounced scattering of points and a less clear overall energy surface, especially in the final part of the curves,

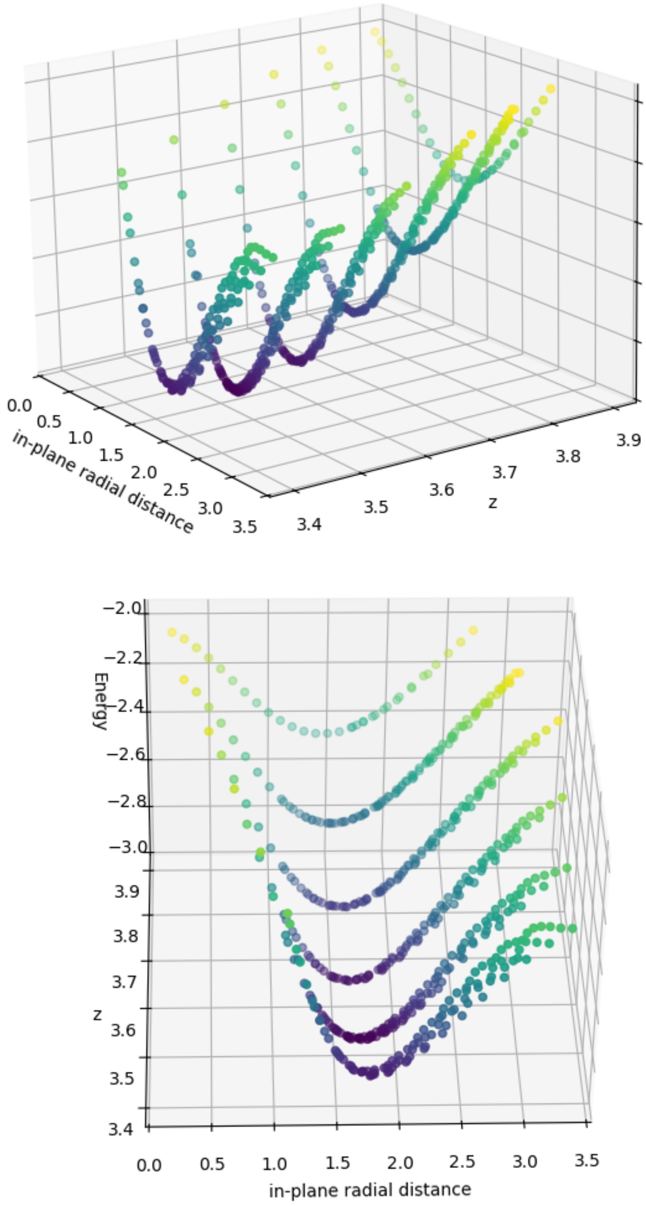


Figure 3.8: Interaction energy as a function of the planar radial distance with vertical distance ≥ 3.4 and energy ≤ -2

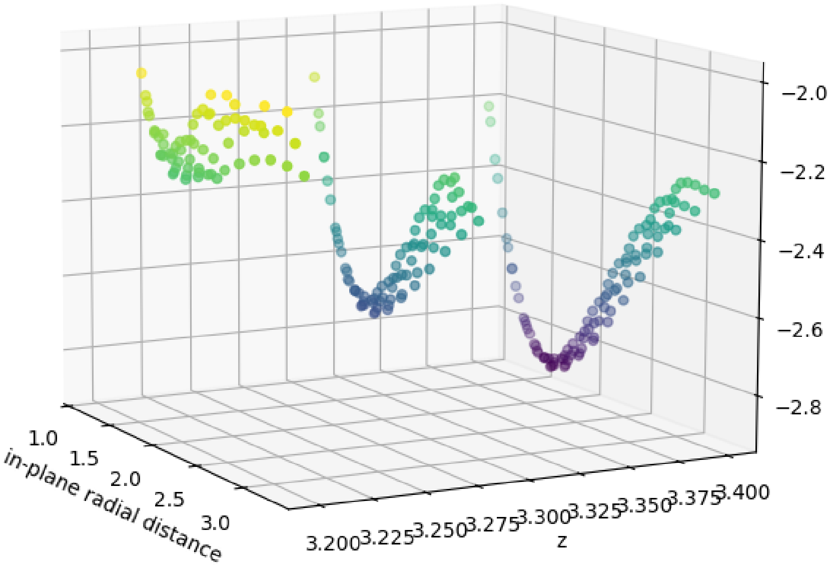


Figure 3.9: Interaction energy for $z = 3.2$, 3.3, and 3.4 Å.

where energy varies less regularly. This behavior is consistent with what was previously highlighted: for smaller vertical distances, the sensitivity of the interaction to spatial variations increases significantly, making the energy surface more complex and less predictable.

We observe that in the equilibrium parallel-displaced configuration, the molecules are translated along a carbon–carbon axis. Therefore, a plausible reason for the planar radial dependence could be related to the fact that, by shifting the molecule along the y-axis for x values smaller than the equilibrium value, one may approach an equilibrium position *along another axis*. Indeed, in the equilibrium configuration—i.e., when the center of the second molecule is located at $(1.7, 0, 3.5)$—the energy value is -2.944 kcal/mol and the radius, introduced in equation 3.2, is, of course, equal to 1.7 Å. On the other hand, when the center of the second molecule is located at $(1.4, 1, 3.5)$, the energy value is -2.942 kcal/mol and the radius is 1.720 Å.

3.3.2 Parallel-displaced: rotations

After obtaining sufficiently representative results regarding spatial translations between the two benzene fragments, we extended the analysis to include rotations, in order to better understand how the interaction energy varies with the relative orientation of the molecules. This in-depth investigation is essential to gaining a comprehensive view of the π-stacking phenomenon, as aromatic interactions are highly sensitive not only to distance but also to the alignment of molecular planes. As a starting point, we chose the equilibrium configuration corresponding to the *parallel-displaced* arrangement, characterized by a horizontal displacement of 1.7 Å, a vertical distance z = 3.5 Å, and zero translation along the y-axis. From this geometry, we performed rotations of the second molecule around the three main axes. In particular, we kept fixed the molecule centered at (1.7, 0.0, 3.5) Å and rotated the one centered at the origin, in order to observe how the interaction energy varies with the orientation of the ligand's benzene fragment relative to the position of the protein's molecule.

It is immediately observed that rotating the molecule within its own plane does not significantly affect the overall interaction energy. To quantitatively analyse this behavior, we initially studied how the energy varies with the rotation angle around the z-axis, i.e., the axis perpendicular to the benzene fragment's plane. Due to the hexagonal symmetry of the benzene molecule, we considered rotation angles between 0° and 30°. It is sufficient to analyse this interval, since a 60° rotation corresponds to perfect overlap with the initial configuration, while larger angles are equivalent to smaller rotations in the opposite direction. For instance, rotating the molecule by 40° is equivalent, by symmetry, to a 20° rotation in the opposite direction. This consideration allows us to reduce the configuration space to be explored, without losing generality in the energetic description of the planar rotation behavior. In Table 3.1 we report the interaction energy values as a function of the in-plane rotation.

Regarding the rotations around the other two main axes, we initially

In-plane rotation	**Energy**
0°	-2.944 kcal/mol
5°	-2.943 kcal/mol
10°	-2.942 kcal/mol
15°	-2.940 kcal/mol
20°	-2.938 kcal/mol
25°	-2.937 kcal/mol
30°	-2.937 kcal/mol

Table 3.1: Interaction energy as a function of the in-plane rotation

analysed the trend of the interaction energy for angles of 0°, 30°, and 60°. However, it was observed that even with a 30° rotation, the interaction energy decreases significantly, compromising the effectiveness of the π-stacking interaction. For this reason, to analyse with greater precision the behavior in the energetically relevant region, we restricted the study to rotations between 0° and 15°. In Table 3.2 (a) we report the results obtained for rotations around the x-axis , followed by those for rotations around the y-axis (3.2 (b)). In the latter case, the rotation lacks symmetry with respect to the initial configuration, so the simulations were conducted for angles between $-15°$ and 15°, with an incremental step of 5°. Even for very small angles, we observe that the energy decreases rapidly with respect to the interaction, highlighting that the parallel component is fundamental for π-stacking interactions. As expected, the rotations of the vertically oriented molecule around the x and y axes cause a significant reduction in the interaction energy. This effect is attributable to the loss of optimal alignment between the two aromatic fragments, which compromises the efficiency of the π-stacking.

In Table 3.2 (c) we report the interaction energy values corresponding to the combined rotations around both axes.

Table 3.2: Interaction energies for different rotations: (a) around the x-axis, (b) around the y-axis, and (c) combined x- and y-axis rotations.

(a)

x-axis rotation	**Energy (kcal/mol)**
0°	-2.944
5°	-2.892
10°	-2.722
15°	-2.409

(b)

y-axis rotation	**Energy (kcal/mol)**
-15°	-2.258
-10°	-2.666
-5°	-2.880
0°	-2.944
5°	-2.887
10°	-2.738
15°	-2.505

(c)

x-axis rotation	**y-axis rotation**	**Energy (kcal/mol)**
5°	5°	-2.840
5°	10°	-2.693
5°	15°	-2.460
10°	5°	-2.687
10°	10°	-2.550
10°	15°	-2.317
15°	5°	-2.408
15°	10°	-2.288
15°	15°	-2.063

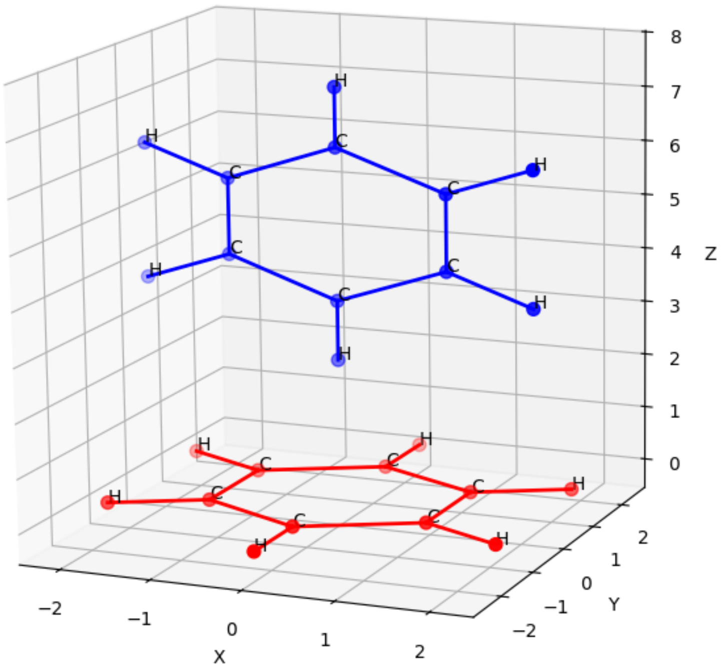


Figure 3.10: Spatial representation of the *T-shaped* configuration used in the simulations.

3.3.3 T-shaped: translations

We now present the results related to the interaction energy in the *T-shaped* configuration. The starting point is the equilibrium configuration, in which one benzene molecule lies on the xy-plane with its center at the origin, while the second molecule is oriented perpendicularly to the plane, with its center in (0, 0, 5).

Starting from this reference geometry, we translated the vertically oriented molecule along the three Cartesian axes, in order to explore how the interaction energy varies with the relative position of the two fragments. To facilitate the understanding of the molecular orientation used in the simulations, we show a graphical representation in Figure 3.10.

It is observed that the energy values associated with the *T-shaped* configuration, at the equilibrium position, are slightly more favorable than those obtained for the *parallel-displaced* configuration. In particular, the minimum energy calculated for the *parallel-displaced* configuration is −2.944 kcal/mol, whereas for the *T-shaped* configuration the minimum value is slightly lower, reaching −3.101 kcal/mol. Although this difference is modest, it highlights that the *T-shaped* configuration is energetically more stable in the context of our simulations.

In Figure 3.11 (a) we show the trend of the interaction energy as a function of the translations along the x and z-axes, while 3.11 (b) we show the plot related to translations along y and z.

It can be seen that the interaction energy is more sensitive to translations along the y-axis than along the x-axis, particularly for small values of the vertical distance z. This behavior can be attributed to the geometric asymmetry introduced by the perpendicular arrangement of the molecules, which makes the lateral component along y more relevant for electronic overlap between the two fragments.

As done previously for the *parallel-displaced* configuration, in Figure 3.12 we report the interaction energy as a function of the vertical distance z and the planar radial distance between the centers of the two molecules. In this representation, only energy values less than or equal to −2 kcal/mol are considered, in order to more clearly highlight the energetically significant configurations and avoid introducing excessive data dispersion.

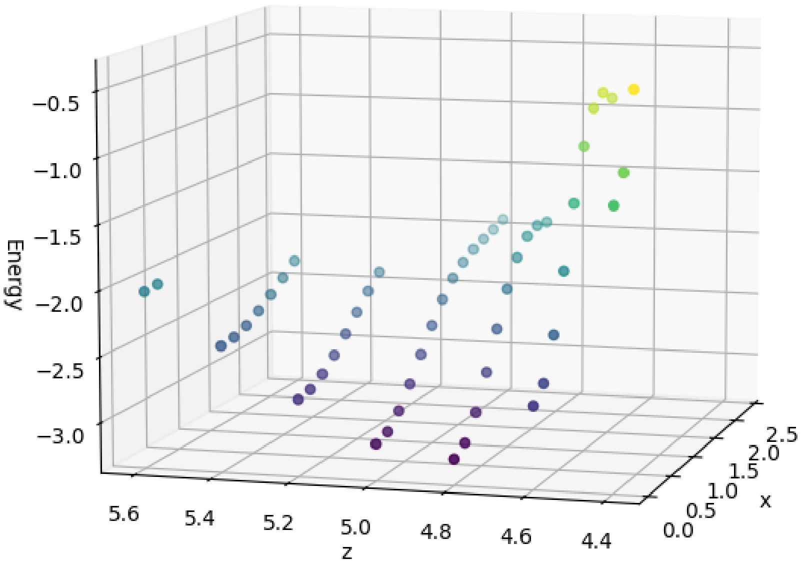
3.3.4 T-shaped: rotations

In Table 3.3 we report the results related to the rotations of the vertically oriented molecule in the *T-shaped* configuration. A first evident aspect is that the total interaction energy remains practically unchanged as the rotation angle around the z-axis varies, i.e., the vertical axis along which the upper molecule is oriented.

The slight differences in the values obtained can likely be attributed to numerical uncertainties inherent to the simulation process, rather than to actual variations in the physical interaction between the molecules.

In Table 3.4 (a) we report the interaction energy values corresponding to

(a)



(b)

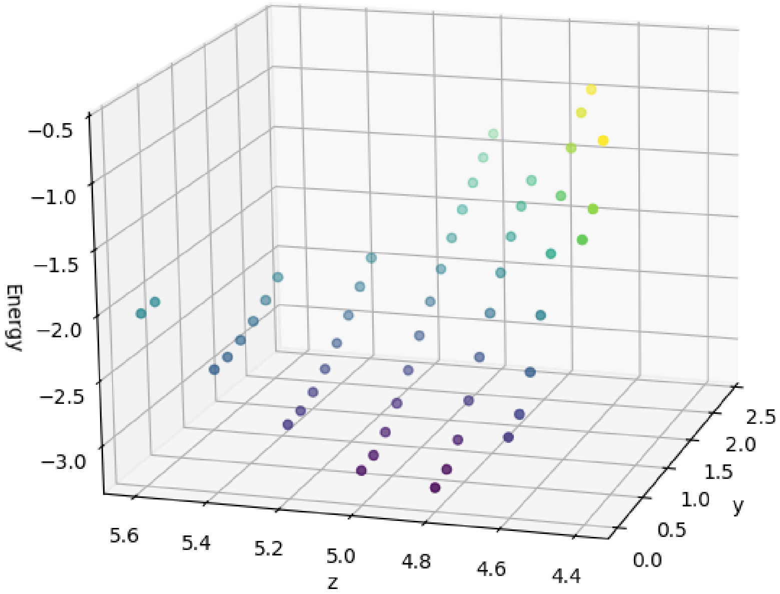


Figure 3.11: Interaction energy as a function of translation along: (a) x and z; (b) y and z.

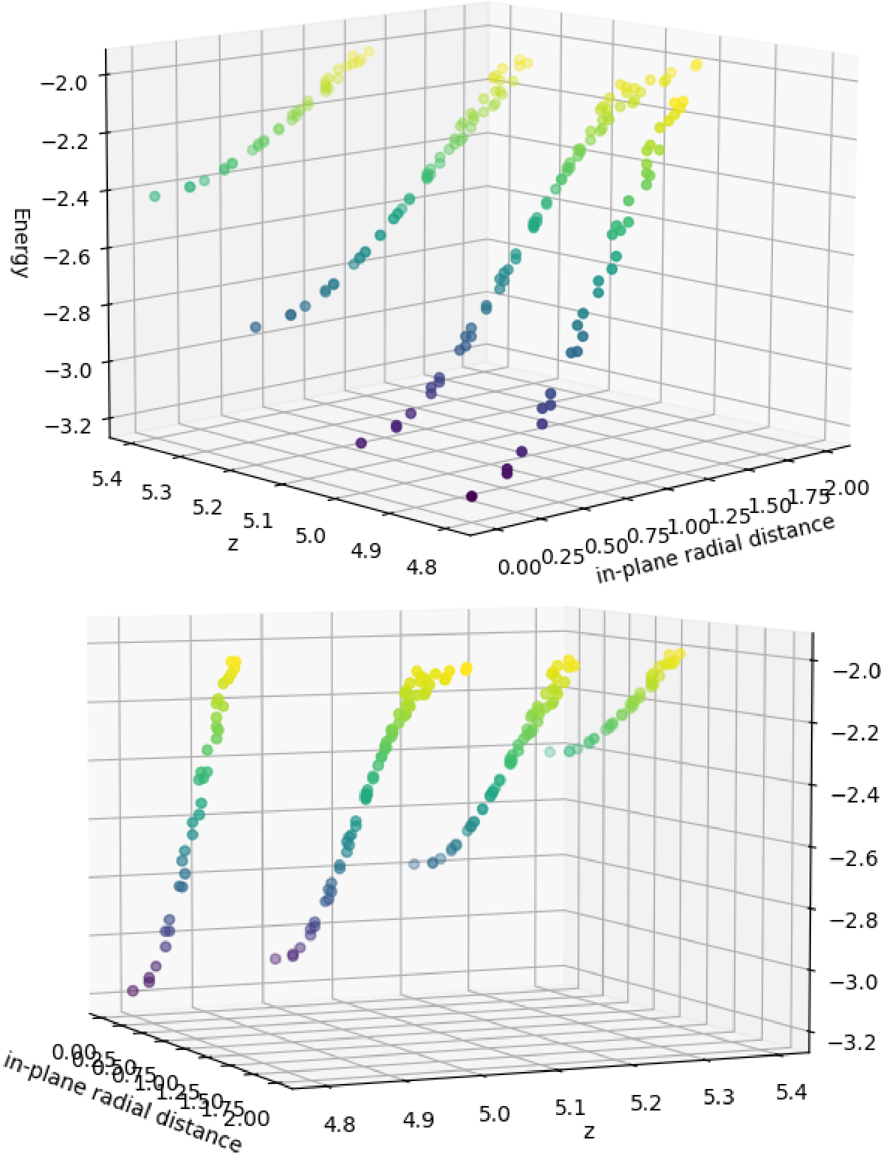


Figure 3.12: interaction energy as a function of the vertical distance z and the planar radial distance, with Energy ≤ -2 kcal/mol

z-axis rotation	**Energy (kcal/mol)**
0°	-3.101
5°	-3.103
10°	-3.103
15°	-3.103
20°	-3.102
25°	-3.102
30°	-3.102

Table 3.3: Interaction energy for different rotations around the z-axis

different rotations around the x-axis, while in 3.2 (b) we show the results for rotations around the y-axis. We observe that the T-shaped configuration is significantly less sensitive to rotations compared to the parallel-displaced configuration. In fact, even with 15° rotations, the intensity of the interaction does not decrease significantly, especially with respect to rotations around the y-axis.

In Table 3.4 (c) we also report the energy values for combined rotations around both the x- and y-axes.

Table 3.4: Interaction energies for different rotations: (a) around the x-axis, (b) around the y-axis, and (c) combined x–y rotations.

(a)

x-axis rotation	**Energy (kcal/mol)**
0°	-3.101
5°	-3.072
10°	-2.980
15°	-2.827

(b)

y-axis rotation	**Energy (kcal/mol)**
0°	-3.101
5°	-3.081
10°	-3.025
15°	-2.940

(c)

x-axis rotation	**y-axis rotation**	**Energy (kcal/mol)**
5°	5°	-3.052
5°	10°	-2.998
5°	15°	-2.918
10°	5°	-2.963
10°	10°	-2.917
10°	15°	-2.854
15°	5°	-2.814
15°	10°	-2.785
15°	15°	-2.742

3.4 Polynomial Fit

In an effort to determine an explicit and analytically manageable functional form for the computed energy surfaces, we decided to perform a systematic fitting procedure. Specifically, we explored both polynomial and rational functions as potential candidates for representing the underlying energy landscape. The goal of this approach is to obtain a continuous mathematical expression capable of accurately reproducing the calculated data points, while remaining simple enough to facilitate further analysis and potential incorporation into computational models. In particular, we chose to accept only those fits for which the coefficient of determination (R^2) exceeded 0.95. This threshold was selected to ensure a high level of accuracy in the representation of the computed energy surfaces.

3.4.1 Parallel-displaced: Displacement Surface Fitting

As a first step, we present a fit of the surface representing the interaction energy as a function of the lateral displacement and the vertical distance between the benzene fragments in parallel-displaced configuration.

After several trials with different functional forms, we opted for third-degree polynomials, which provided a sufficiently accurate approximation of the surface while maintaining a reasonable level of complexity. Higher-degree polynomials or more elaborate functional forms were deliberately avoided to preserve interpretability.

We restricted our analysis to configurations where z values lie between 3.3 Å and 4 Å, and where the interaction energy is less than or equal to -2 kcal/mol. This selection criterion was adopted to focus on the most physically relevant region of the energy landscape, thereby avoiding the influence of extreme or non-representative data points. By doing so, we aimed to produce a fit that is both effective and meaningful for the energy range of interest.

For illustrative purposes, in Figure 3.13 we also present the result obtained using a second-degree polynomial, which, despite its simplicity, offers a qualitatively reasonable description of the energy surface.

The quadratic polynomial obtained from the fitting procedure is:

$$E(d,z) = 0.412\,d^2 + 3.279\,z^2 + 0.785\,dz - 4.296\,d - 24.797\,z + 45.270,$$

where d represents the lateral displacement and z the vertical distance between the benzene fragments. The coefficient of determination (R^2) for this fit is 0.912, indicating a good correlation between the fitted surface and the computed data points. However, it is important to note that the minima of the energy surface are not accurately captured by this quadratic fit. For this reason, we proceeded to perform a third-degree polynomial fitting, which is presented in Figure 3.14, in an attempt to achieve a more accurate representation of the energy landscape, particularly in the regions close to the minima. The third-degree polynomial obtained from the fitting procedure is:

$$\begin{aligned} E(d,z) = {} & -0.092\,d^3 - 3.951\,z^3 - 1.001\,d^2z - 1.563\,dz^2 + 4.601\,d^2 \\ & + 49.543\,z^2 + 15.848\,dz - 39.418\,d - 207.327\,z + 287.824. \end{aligned}$$

The coefficient of determination (R^2) for this cubic fit is 0.964, indicating a significant improvement in the accuracy of the surface approximation, especially in the regions near the energy minima.

3.4.2 Parallel-displaced: Radial Surface Fitting

In Figure 3.15, we present the fitting results for the interaction energy as a function of both the radial distance and the vertical separation between the planes of the benzene fragments. Also in this case, restricted our analysis to configurations where z values lie between 3.2 Å and 4 Å, and where the interaction energy is less than or equal to -2 kcal/mol. The third-degree polynomial obtained from the fitting procedure is:

$$\begin{aligned} E(d,z) = {} & -0.178\,d^3 - 4.544\,z^3 - 0.905\,d^2z - 1.432\,dz^2 + 4.688\,d^2 \\ & + 55.842\,z^2 + 14.549\,dz - 37.063\,d - 229.286\,z + 313.113. \end{aligned}$$

The R^2 coefficient for this cubic fit is equal to 0.956.

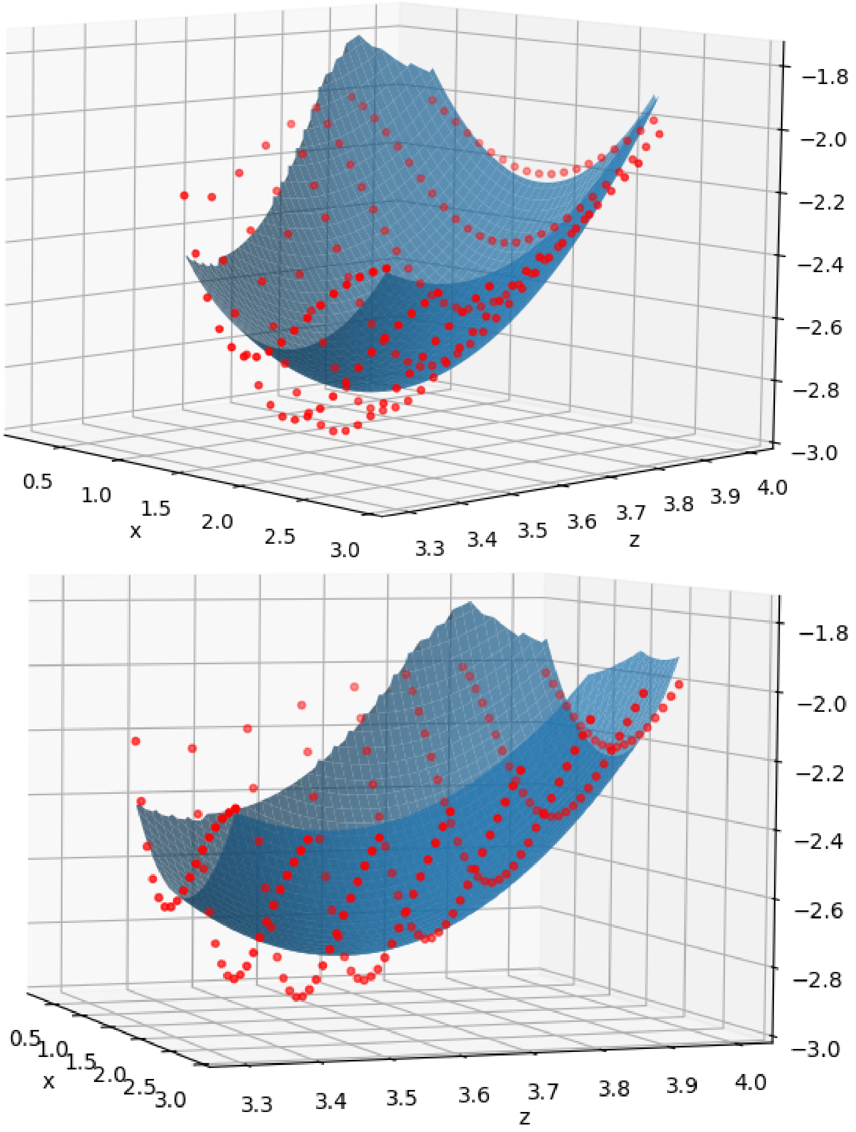


Figure 3.13: Quadratic interpolation of the interaction energy as a function of lateral displacement and vertical separation

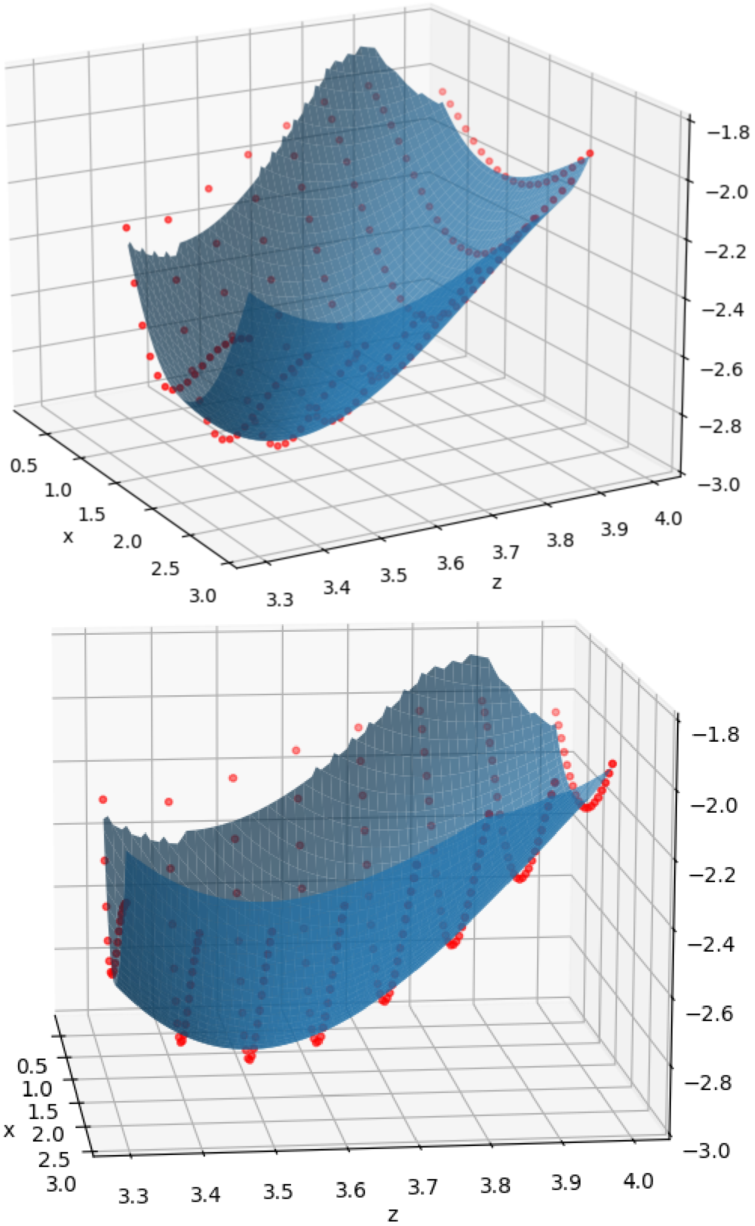


Figure 3.14: Cubic interpolation of the interaction energy as a function of lateral displacement and vertical separation

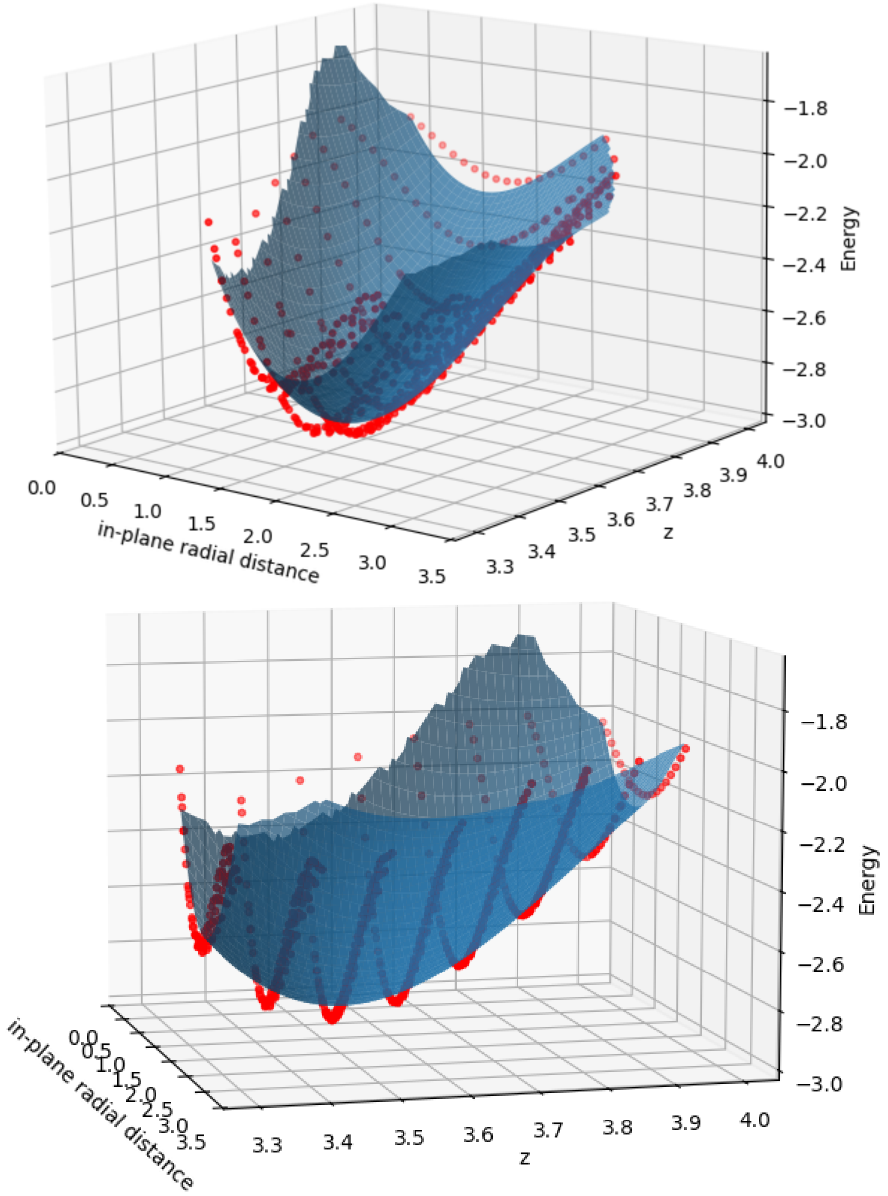


Figure 3.15: Cubic interpolation of the interaction energy as a function of radial distance and vertical separation

3.4.3 T-shaped: x and y displacement fits

In Figure 3.16 (a) we present the fitting of the energy surface corresponding to the T-shaped configuration, where the interaction energy is expressed as a function of the displacement along the x-axis and the vertical distance between the centers of the benzene fragments.

The third-degree polynomial obtained from the fitting of the T-shaped configuration is:

$$\begin{aligned} E(d,z) = & -0.331\,d^3 - 3.632\,z^3 - 0.209\,d^2z + 1.106\,dz^2 + 1.922\,d^2 \\ & + \ 57.868\,z^2 - 12.294\,dz + 33.943\,d - 305.206\,z + 530.162. \end{aligned}$$

The coefficient of determination (R^2) for this cubic fit is 0.995, indicating an excellent agreement between the fitted surface and the computed energy data for the T-shaped configuration.

In Figure 3.16 (b), instead, we represent the interaction energy as a function of the displacement along the y-axis and the vertical distance z between the centers of the benzene fragments.

The interpolating third-degree polynomial for this surface is:

$$\begin{aligned} E(d,z) = & -0.269\,d^3 - 4.002\,z^3 - 0.335\,d^2z + 1.116\,dz^2 + 2.563\,d^2 \\ & + \ 63.013\,z^2 - 12.041\,dz + 32.316\,d - 328.813\,z + 565.826. \end{aligned}$$

The coefficient of determination (R^2) for this cubic fit is 0.999, confirming an excellent level of agreement between the interpolating polynomial and the computed energy values.

3.4.4 T-shaped: Radial Surface Fitting

In Figure 3.17 we present the fitting of the surface obtained by plotting the interaction energy as a function of the radial distance within the plane and the vertical separation between the centers of the benzene rings for the T-shaped configuration.

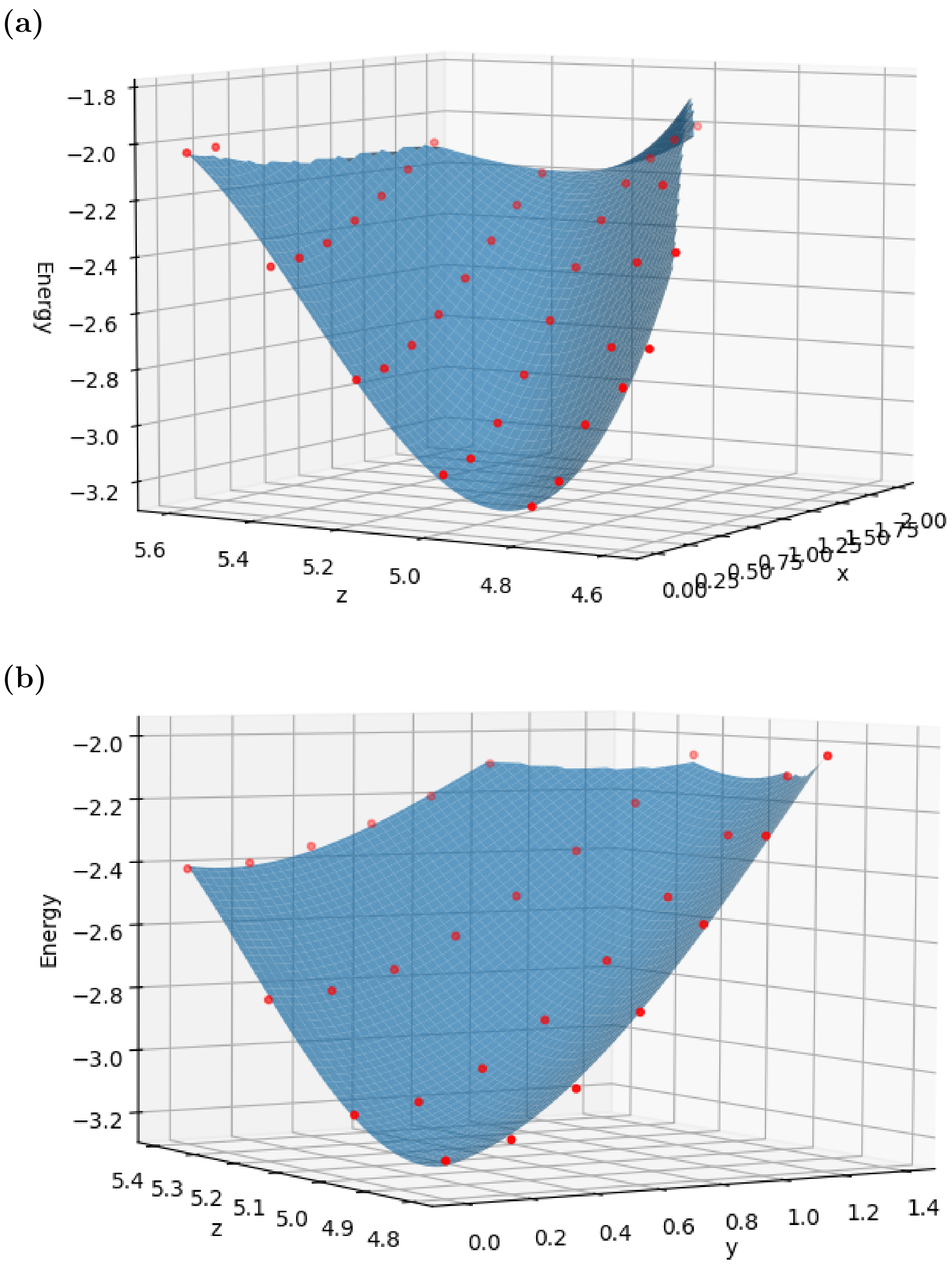


Figure 3.16: Interpolation of the interaction energy as a function of the displacement along the (a) x-axis; (b) y-axis.

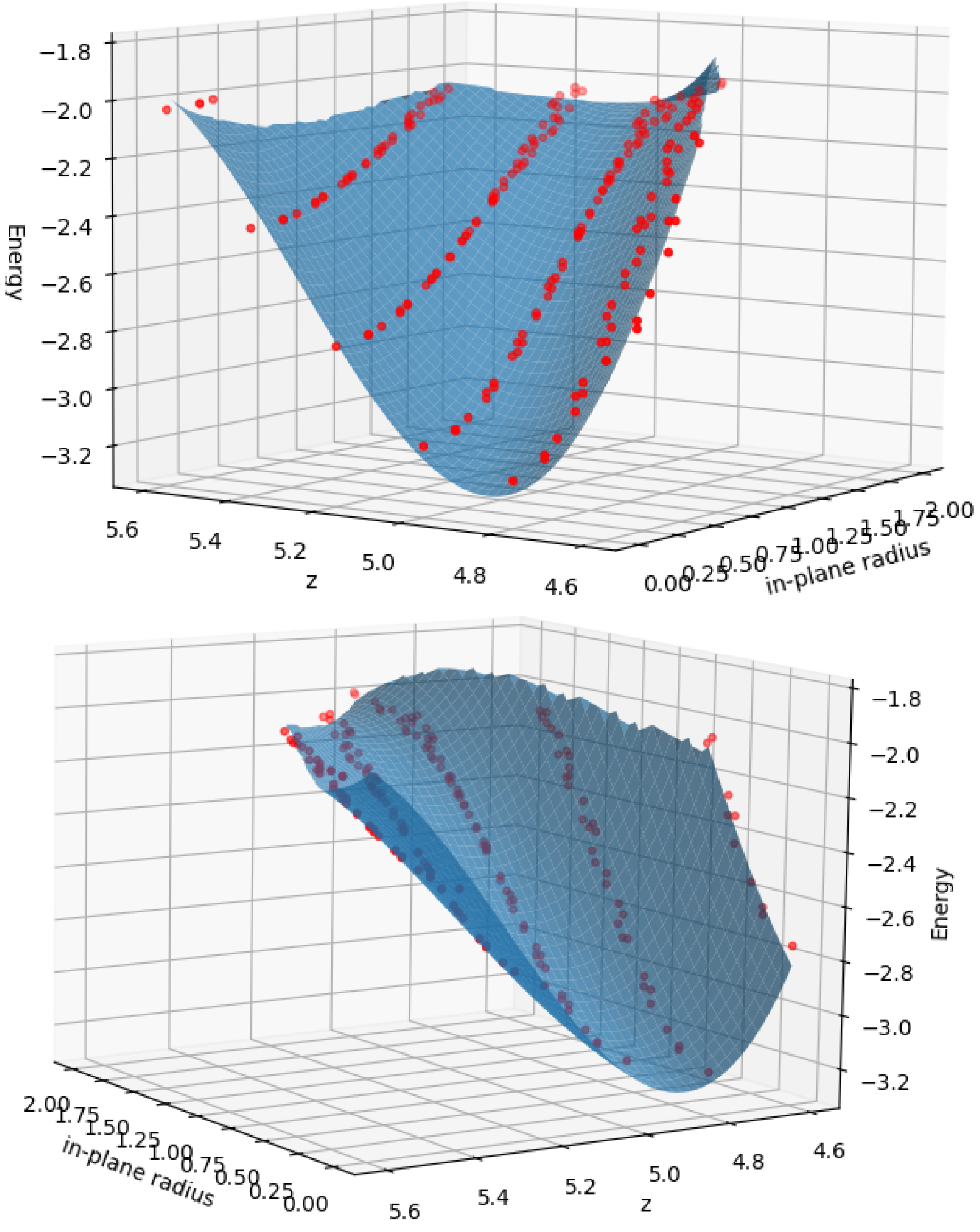


Figure 3.17: Interaction energy as a function of the radial distance within the plane and the vertical separation

The third-degree polynomial corresponding to this surface is:

$$\begin{aligned}E(d,z) = &-0.347\,d^3 - 4.060\,z^3 + 0.151\,d^2z + 0.748\,dz^2 + 0.168\,d^2\\ &+\ 64.453\,z^2 - 9.071\,dz + 26.761\,d - 338.774\,z + 586.878.\end{aligned}$$

The coefficient of determination (R^2) for this cubic fit is 0.990, indicating a very good agreement between the polynomial model and the computed energy data points.

Chapter 4

π-stacking mathematical modelling

Our goal is to understand and define how to integrate the π-stacking interaction into the physical Hamiltonian introduced in Section 2.1.3. This integration presents a specific challenge: introducing an interaction of molecular nature into a model designed to naturally handle atomic-level interactions.

In fact, the use of a graph to represent the positions of the atoms in the protein and the possible positions occupied by the ligand's atoms makes the model naturally suited to describe interatomic forces - such as electrostatic or Van der Waals interactions - but not intuitive for representing molecular-level interactions. For this reason, several steps were necessary to achieve a consistent modeling.

4.1 Identification of Benzenic Fragments in the Protein

The first step is to identify the benzenic fragments of the protein that are sufficiently close to the binding pocket. This allows us to determine which benzene molecules in the protein can actually be subject to a π-stacking interaction, in case another benzenic fragment is placed in the pocket. The

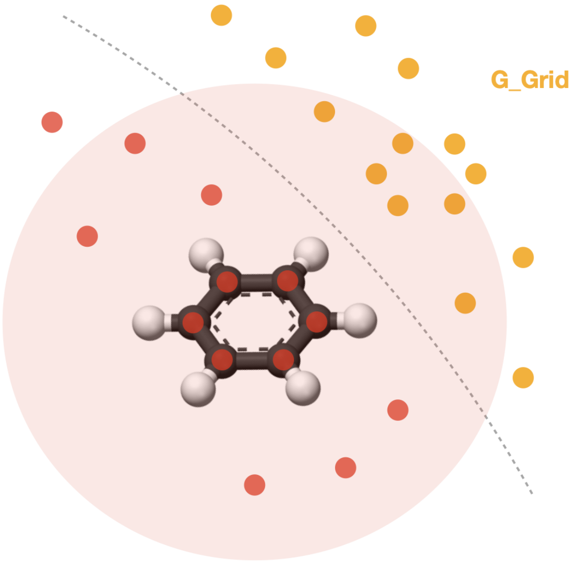


Figure 4.1: Selecting the benzene fragment near the pocketgrid

idea behind this approach is that if a benzene molecule is far from the pocket, the π-stacking interaction will be negligible.

To identify such molecules, the following procedure is applied: all benzenic fragments within the protein are identified, the center C of the aromatic ring is determined, and a sphere of radius r centered at C is constructed. A benzenic fragment is considered relevant if the sphere intersects at least one node of the pocket graph G_{grid}, as presented in Figure 4.1. Otherwise, the fragment is disregarded.

After identifying, through the previous analysis, the benzenic fragments of the protein potentially involved in π-stacking interactions, the next step is to identify the vertices of the graph G_{grid} that could be affected by such interaction. In other words, we are interested in determining which nodes of the pocketgrid (4.2) are sufficiently close to a benzenic fragment of the protein, such that if a benzene molecule from the ligand were placed there, a significant π-stacking interaction would occur.

In the following discussion, we focus on the case of the *parallel displaced* interaction, noting that the modeling for the *T-shaped* configuration follows

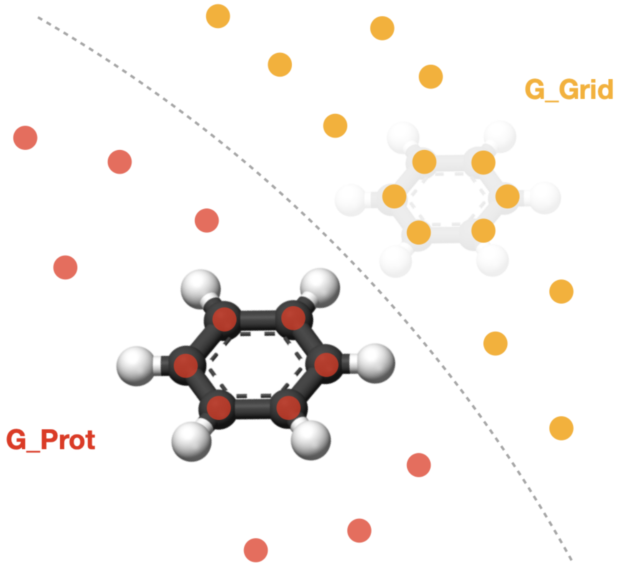


Figure 4.2: Nodes of the pocketgrid sufficiently close to a benzenic fragment

an entirely analogous procedure.

4.2 First Modeling Approach

Let $\mathscr{C} = \{C^{(1)}, C^{(2)}, \ldots, C^{(6)}\}$ be a benzenic fragment in the protein identified as potentially interacting with the pocket. The goal is to determine a collection of sets of six vertices $\mathscr{D} = \{D^{(1)}, D^{(2)}, \ldots, D^{(6)}\}$, such that, if a benzene molecule from the ligand were placed at those vertices, a significant energy contribution due to the π-stacking interaction would be generated.

Since the graph G_{grid} is a discretization of three-dimensional space, the ligand molecule cannot be placed in an arbitrary position, but is constrained to the vertices of the graph. It is therefore essential to determine not only which nodes can give rise to a π-stacking interaction, but also how much they deviate from the ideal position, that is, the one that minimizes the interaction energy.

To construct the set $\mathscr{D}$, we adopt a vector-based approach. Let C denote the center of the benzenic ring $\mathscr{C}$, computed as the midpoint between two

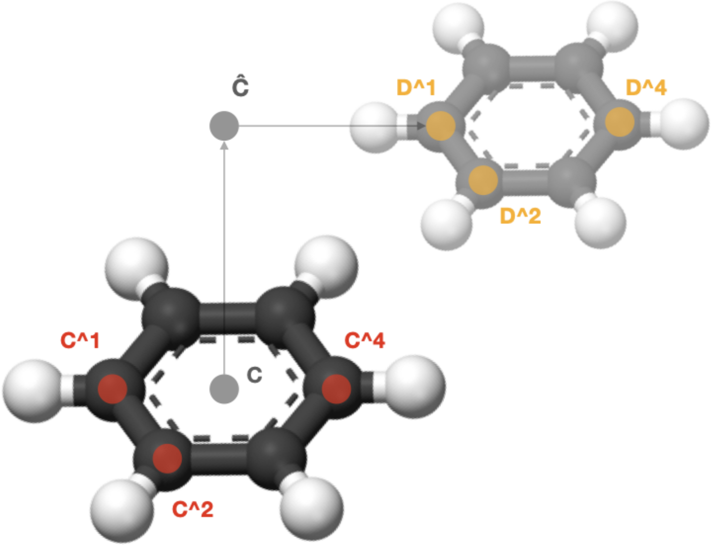


Figure 4.3: Graphical representation of $\mathscr{C}$ and $\mathscr{D}$

opposite carbon atoms:

$$C = C^{(1)} + \tfrac{1}{2}\big(C^{(4)} - C^{(1)}\big)\,.$$

We then define a vector $\hat{C}$, orthogonal to the plane of the ring, obtained by normalizing the cross product of two in-plane vectors and multiplying by the equilibrium vertical distance:

$$\hat{C} = \pm 3.5\,\frac{(C^{(1)} - C) \times (C^{(2)} - C)}{\big\|(C^{(1)} - C) \times (C^{(2)} - C)\big\|}\,,$$

where the sign ± is chosen according to the desired orientation.

Based on this, for each vertex $C^{(i)}$, we construct a corresponding point $D_i^{(1)}$ lying in the translated plane of the second molecule, defined as:

$$D_i^{(1)} = \hat{C} + \frac{1.7}{1.4}\,(C^{(i)} - C)\,.$$

Since the nodes of G_{grid} do not cover space continuously, it is unlikely that a vertex exists exactly at each point $D_i^{(1)}$. For each point $D_i^{(1)}$, we therefore consider a sphere of radius $\hat{r}$ centered at $D_i^{(1)}$ and check whether it contains at least one node of the graph. If such a node exists, it is selected as a candidate and the search proceeds for the other five corresponding vertices, following the pattern of the benzenic molecule.

Now, suppose a vertex has been identified near $D_i^{(1)}$. We then proceed to determine the remaining five vertices of the hexagon. In particular, the point $D^{(4)}$ is searched as the intersection between G_{grid} and a sphere centered at $D^{(1)} + (C^{(4)} - C^{(1)})$. Then, a plausible arc is defined to search for $D^{(2)}$, and the same criterion is applied to find the remaining vertices.

Once three non-aligned vertices are identified, the position of the remaining ones is uniquely determined. This is because three points in space define a plane, and the hexagonal structure of benzene imposes fixed distances between atoms.

At this point, it is possible to assess how much the obtained configuration deviates from the ideal one. Recall that the optimal conditions for the π-stacking interaction require that the second molecule:

- is parallel to the first molecule;
- is rotated with respect to the first by an angle equal to a multiple of 60°;
- is located at a vertical distance of 3.5 Å;
- has a horizontal distance from the center of the first molecule equal to 1.7 Å.

Let D denote the geometric center of the set $\mathscr{D}$. We then verify the following geometric conditions:

- Orthogonality with respect to the vertical direction:

$$(D^{(4)} - D^{(1)}) \cdot (\hat{C} - C) = 0, \quad (D^{(2)} - D^{(1)}) \cdot (\hat{C} - C) = 0;$$

- Orthogonality between corresponding sides:

$$(C^{(2)} - C^{(1)}) \cdot (D^{(2)} - D^{(1)}) = 0;$$

- Horizontal distance:

$$\frac{\left|(D - C) \cdot (C^{(4)} - C)\right|}{\|C^{(4)} - C\|} = 1.7;$$

- Vertical distance:

$$\frac{\left|(D - C) \cdot (\hat{C} - C)\right|}{\|\hat{C} - C\|} = 3.5;$$

- Alignment between the benzene orientation and the horizontal axis:

$$\frac{(D^{(1)} - \hat{C}) \cdot (C^{(4)} - C)}{\left|(D^{(1)} - \hat{C}) \cdot (C^{(4)} - C)\right|} = 1.$$

For each of the conditions above, we can estimate a numerical value indicating how well it is satisfied. Based on these measures, we define an intensity coefficient $I \in [0, 1]$, which quantifies the degree of alignment between the actual configuration and the ideal one. The energy associated with such an interaction is then given by:

$$E = -2.94 \cdot I\,.$$

It is important to highlight that the value -2.94 represents the minimum interaction energy obtained from the simulations discussed in Section 3.2. This value corresponds to the total energetic contribution associated with the interaction between aromatic rings in a specific configuration and therefore includes multiple components of different physical nature. In this chapter, the Hamiltonian will refer to the total interaction energy. However, in Chapter 5, we will instead focus on a *residual* molecular contribution — an effective interaction term that excludes the main atomistic components, such as Coulomb and Van der Waals forces.

In the case where, for a given benzenic fragment of the protein, multiple plausible configurations are identified (i.e., multiple sets of six nodes in G_{grid}), all of them must be considered, introducing one variable I for each.

To enrich the molecular representation within the protein graph G_{mol}, we introduce new labels useful for identifying benzenic fragments. Specifically, to each node i we assign a label c_i, defined as:

$$c_i = \begin{cases} 1, & \text{if } i \text{ is a carbon atom belonging to a benzenic fragment,} \\ 0, & \text{otherwise.} \end{cases}$$

In addition, we associate to node i an ordered pair $\{i_1, i_2\}$, which identifies the two carbon atoms adjacent to i in the benzenic ring. If i does not belong to any benzenic ring, we conventionally set $i_1 = i_2 = i$.

The graph G_{mol} is thus extended as follows:

$$G_{\mathrm{mol}} = \{i,\ e_{ij},\ w_{ij},\ q_i,\ c_i,\ \{i_1, i_2\}\}\,.$$

Similarly, the graph G_{grid}, which represents the spatial discretisation of the protein pocket, is also enriched with new information.

To each node i', we assign a label $I_{i'}$, which represents the intensity of the interaction in the case where i' coincides with one of the nodes $D_i^{(1)}$ defined previously. If the node i' does not belong to any set $\mathscr{D}$, we simply set $I_{i'} = 0$.

As in the previous case, we also associate a pair $\{i'_1, i'_2\}$ to the node i', according to the same criterion adopted for G_{mol}, to maintain structural consistency between the two graphs.

The extended graph G_{grid} thus takes the form:

$$G_{\mathrm{grid}} = \{i',\ e_{i'j'},\ w_{i'j'},\ V_{i'},\ I_{i'},\ \{i'_1, i'_2\}\}\,. \tag{4.1}$$

We now introduce a variable $\lambda_{ii'}$, defined as the product between the indicator of belonging to a benzenic ring c_i and the interaction intensity $I_{i'}$ at node i':

$$\lambda_{ii'} := c_i I_{i'}.$$

This expression takes a positive value only when node i belongs to a benzene molecule and i' is a node in the graph G_{grid} that experiences a significant π-stacking interaction. In all other cases, $\lambda_{ii'} = 0$, and the associated energy contribution is null.

Based on the definition of $\lambda_{ii'}$, we express the Hamiltonian term associated with the π-stacking interaction in Equation 4.2:

$$H_\pi = \sum_{i,i'} (-2.94)\, \lambda_{ii'}\, x_{ii'} \left(x_{i_1 i'_1 i_2 i'_2} + x_{i_2 i'_1 i_1 i'_2} \right), \tag{4.2}$$

where:

- the binary variable $x_{ii'}$ equals 1 if node i in the graph G_{mol} is mapped to node i' in the graph G_{grid}, and 0 otherwise;
- the variable $x_{i_1 i'_1 i_2 i'_2}$ equals 1 if i_1 is mapped to i'_1 and simultaneously i_2 is mapped to i'_2. Similarly, $x_{i_2 i'_1 i_1 i'_2}$ accounts for possible 180-degree rotations in the ligand.

This expression contributes negatively to the system's energy only when the correspondence between nodes respects the local structure of benzene, thus favoring alignments consistent with a π-stacking interaction. The term H_π therefore describes the negative (favorable) energy contribution that arises when a carbon atom belonging to a benzenic ring is correctly positioned in a region of the pocket where a π-stacking interaction is observed, and its adjacent atoms are also positioned coherently with the benzene structure.

In the case where the graph G_{grid} contains only a very limited number of hexagonal configurations compatible with the π-stacking interaction, the problem is relatively easy to handle - in particular, if there is an almost one-to-one correspondence between each carbon atom in the protein's benzenic fragments and a possible position for the second benzene molecule in the ligand. However, when many potentially compatible configurations are present in the pocket, it becomes necessary to adopt a more complex approach. This occurs, for example, when around a single node of the graph G_{grid}, a potential candidate for a vertex $D^{(1)}$, multiple sets of six nodes forming valid hexagons are identified, as shown in Figure 4.4.

To properly handle this situation, we introduce a new labeling system for the nodes of the graph G_{grid} [4.3]. Specifically, to each node i' we assign:

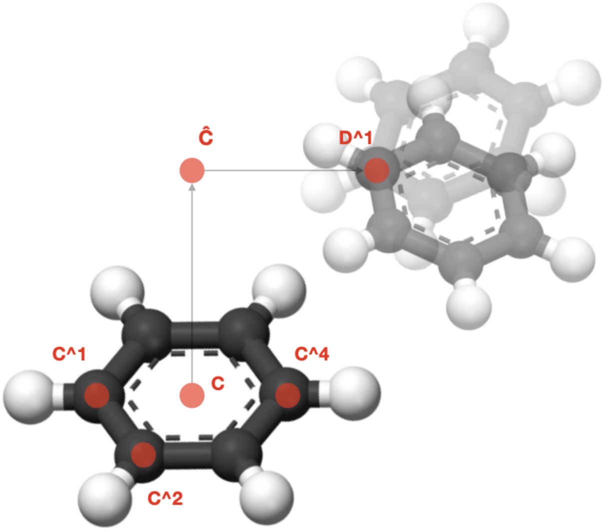


Figure 4.4: Example of a grid region where multiple hexagonal structures can be identified around a single node D^1.

- an integer $l_{i'}$, equal to twice the number of hexagons that include the node i' as a possible vertex $D^{(1)}$;
- an ordered list $\{i'_1, i'_2, i'_3, \ldots, i'_{l_{i'}}\}$, in which each pair $\{i'_{2k-1}, i'_{2k}\}$ represents the two vertices adjacent to i' in one of the compatible hexagons;
- a set of values $I_{i'i'_{2k-1}i'_{2k}}$, which measures the intensity of the interaction corresponding to each identified configuration.

If the node i' does not participate in any hexagonal configuration, we conventionally set $l_{i'} = 1$ and $i'_1 = i'_2 = i'$, in order to maintain a uniform graph structure.

The extended graph G_{grid} is thus described by the following structure:

$$G_{\text{grid}} = \left\{ i',\ e_{i'j'},\ w_{i'j'},\ V_{i'},\ l_{i'},\ \{i'_1, i'_2, \ldots, i'_{l_{i'}}\},\ \{I_{i'i'_1i'_2}, \ldots, I_{i'i'_{l_{i'}-1}i'_{l_{i'}}}\} \right\}. \tag{4.3}$$

To correctly include the energy contribution due to multiple possible local configurations, we also extend the definition of the Hamiltonian H_π. We introduce the term:

$$\lambda_{ii'i'_{2k-1}i'_{2k}} := c_i \cdot I_{i'i'_{2k-1}i'_{2k}}, \quad \text{for } k \in \left\{1, 2, \ldots, \tfrac{l_{i'}}{2}\right\}, \tag{4.4}$$

where c_i represents, as before, the indicator that node i belongs to a benzenic ring.

The general form of the interaction term becomes:

$$H_\pi = \sum_{i,i'} \sum_{k=1}^{\frac{l_{i'}}{2}} (-2.94)\, \lambda_{ii'i'_{2k-1}i'_{2k}}\, x_{ii'} \left(x_{i_1 i'_{2k-1} i_2 i'_{2k}} + x_{i_2 i'_{2k-1} i_1 i'_{2k}} \right), \tag{4.5}$$

where the binary variables $x_{ii'}$ and $x_{i_1 i'_{2k-1} i_2 i'_{2k}}$ have the same meanings as previously defined, with the only difference being that we now consider, for each node, all possible combinations consistent with the presence of multiple local hexagons. This formalism allows for an accurate modeling even in complex cases where the geometry of the pocket features a significant variety of local configurations compatible with π-stacking interactions.

4.3 Second Modeling Approach

At this point, given the impossibility of accurately representing the exact spatial position of the molecule, we opted for an alternative approach: once a benzene molecule in the protein sufficiently close to the pocket is fixed, we select certain regions of space G_{grid} within which, if a benzenic fragment of the ligand were positioned, a π-stacking interaction would occur.

Within these regions — which will be more precisely defined later — we consider a constant average energy. In other words, if a benzenic fragment of the ligand falls into one of these predefined zones, we assume that an average energy contribution is added to the total Hamiltonian of the system.

Specifically, for each benzenic fragment of the protein, we define two new subgraphs of G_{grid}, G_{center} and G_{benzene}.

- G_{center} represents the set of points in G_{grid} which, if occupied by the center of an aromatic ring, would give rise to a π-stacking interaction, assuming the molecule is in a favorable configuration and satisfies the previously imposed constraints.

- G_{benzene} is the set of points in G_{grid} corresponding to the positions of the carbon atoms of the benzenic fragment, in the case where its center is located at a point of G_{center}. In other words, G_{benzene} is a graph that includes G_{center} and is used to fix the spatial arrangement of the carbon atoms of the benzenic fragment.

In Figure 4.5, we show the subgraphs G_{center} and G_{benzene} related to a benzenic fragment of the protein. Specifically, in this case, we show only the region involved in the π-stacking interaction along a particular carbon–carbon axis of the aromatic ring of the protein. Given the symmetry of the problem, the two graphs will ultimately be represented by two tori, both having the same central circumference and different radii.

We introduced new labels for both the molecular graph G_{mol} and the spatial grid G_{grid}, with the aim of explicitly modeling the presence and spatial arrangement of benzenic fragments involved in the π-stacking interaction. For G_{mol}, we associate the following labels with each node i:

- a variable $\hat{c}_i$, which is equal to 1 if node i represents the center of a benzenic fragment, and 0 otherwise;
- a variable c_{ij}, which is equal to 1 if node i is a carbon atom belonging to the benzenic fragment centered at node j, and 0 otherwise.

The enriched structure of the molecular graph is thus:

$$G_{\text{mol}} = \{i,\, e_{ij},\, w_{ij},\, q_i,\, \hat{c}_i,\, c_{ij}\}\,. \tag{4.6}$$

Similarly, for the grid G_{grid}, we introduce:

- a label $\hat{c}_{i'}$, which is equal to 1 if vertex i' belongs to the subgraph G_{center}, i.e., if it can represent the center of a benzenic fragment, and 0 otherwise;

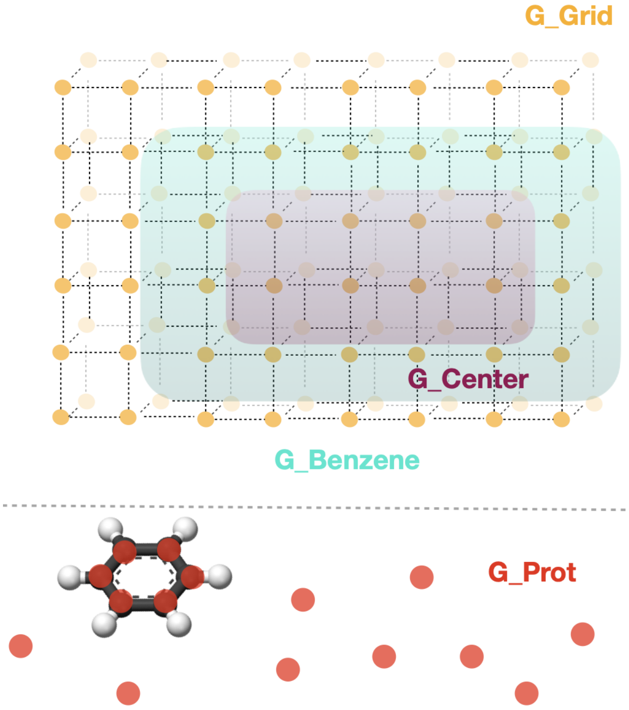


Figure 4.5: Graphical representation of G_{center} and G_{benzene}, corresponding to the benzenic fragment within the protein structure.

- a variable $c_{i'j'}$, which is equal to 1 if vertex i' belongs to the subgraph G_{benzene} and j' belongs to the corresponding G_{center} subgraph, and 0 otherwise.

The structure of the grid graph is therefore:

$$G_{\text{grid}} = \{i',\, e_{i'j'},\, w_{i'j'},\, V_{i'},\, \hat{c}_{i'},\, c_{i'j'}\}\,. \tag{4.7}$$

Based on these labels, we can finally define the contribution to the total Hamiltonian, via the term H_π, given by:

$$H_\pi = \sum_{i,i'} \sum_{j,j'} \frac{\hat{E}}{6}\, \hat{c}_i\, \hat{c}_{i'}\, c_{ji}\, c_{j'i'}\, x_{ii'}\, x_{jj'}\,, \tag{4.8}$$

where $\hat{E}$ represents a constant average energy value associated with the benzene-benzene interaction.

We now rewrite Equation 4.8 in a clearer form and provide a more detailed analysis:

$$H_\pi = \frac{\hat{E}}{6} \sum_{i,i'} \hat{c}_i \, \hat{c}_{i'} x_{ii'} \, \big(\sum_{j,j'} c_{ji} \, c_{j'i'} \, x_{jj'} \big). \tag{4.9}$$

We observe that, in the first summation, the term $\hat{c}_i \, \hat{c}_{i'} \, x_{ii'}$ takes the value 1 if and only if the center of a benzenic fragment i is positioned on a node i' belonging to the subgraph G_{center}. This occurs because $\hat{c}_i = 1$ indicates that i is indeed the center of a benzenic fragment, while $\hat{c}_{i'} = 1$ implies that node i' of the grid belongs to the subgraph G_{center}. The second summation contributes to the Hamiltonian by adding a term equal to 1 every time a carbon atom j, belonging to a benzenic fragment centered at i, is assigned to a node j' of the grid that belongs to the subgraph G_{benzene} associated with node $i' \in G_{\text{center}}$. In this way, we account for the full presence of the ligand's benzenic fragment in a spatial region that allows for π-stacking interaction.

Note that, if all six carbon atoms of a benzenic fragment are positioned within the subgraph G_{benzene}, the total contribution of the second summation is equal to 6. For this reason, the Hamiltonian is multiplied by a normalization factor of 1/6, to ensure that the total energy associated with a fully inserted fragment in the interaction region equals the desired average energy value. This energy value is given by the coefficient $\hat{E}$, which represents the average interaction energy when the ligand's benzenic fragment is located in a favorable interaction region.

In Figure 4.6, we show a graphical representation of a benzenic ring satisfying the imposed conditions. Specifically, the benzene on the lower plane represents a benzenic fragment of the protein, while the one on the upper plane is a benzenic fragment of the ligand. The red spatial region represents G_{center}, while the yellow region represents G_{benzene}.

However, this modeling presents several limitations. The first concerns the lack of control over the molecule's rotation within its own plane. The second relates to the fact that, if the molecule undergoes a generic rotation, it may happen that only part of its atoms lie within the subgraph G_{benzene},

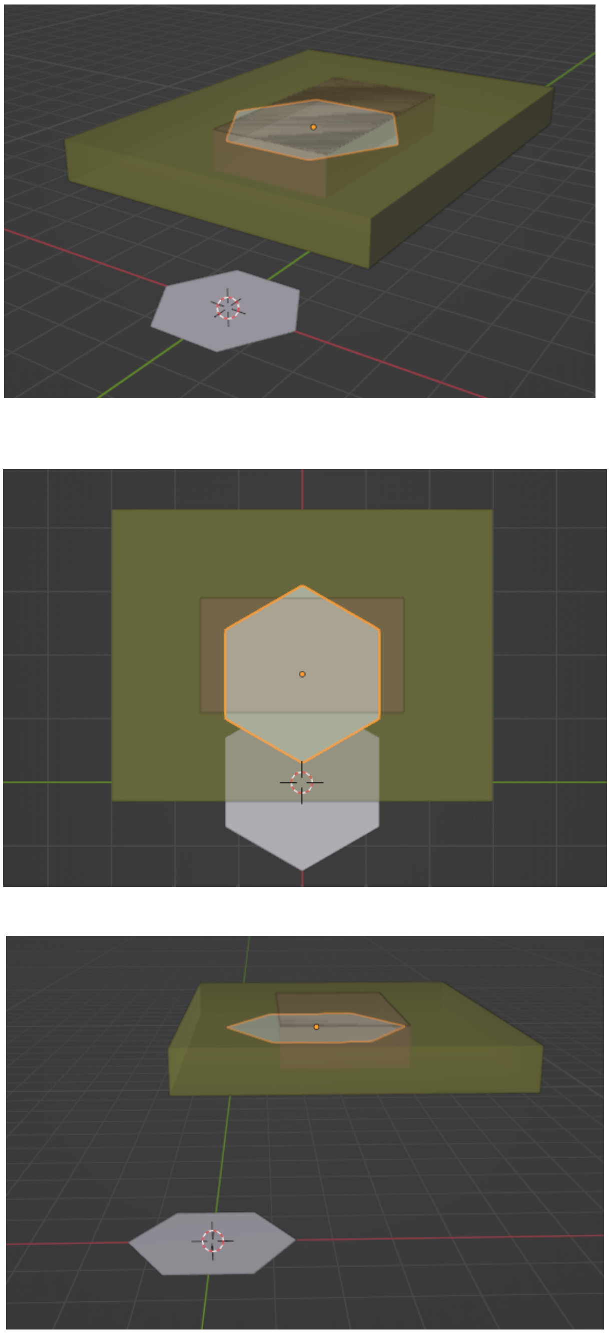


Figure 4.6: Benzenic fragment of the ligand represented within G_{center} and G_{benzene}.

while others remain outside.

As a consequence, the contribution provided by the second summation in Equation 4.9, namely

$$\sum_{j,j'} c_{ji}\, c_{j'i'}\, x_{jj'},$$

may range between 0 and 6, but it does not correctly reflect the actual quality of the interaction. In particular, spatial configurations where the π-stacking interaction should be negligible can still produce a positive energy contribution, leading to a distortion in the overall evaluation of the Hamiltonian.

From the results obtained through numerical simulations presented in Section 3.3.2, we observe that the rotation of the molecule within its own plane has a relatively limited effect on the overall energy of the π-stacking interaction. In contrast, rotations around the other two spatial axes - that is, those altering the orientation of the molecule's plane relative to the interaction surface - lead to a significant reduction in interaction energy.

To account for this behavior and appropriately penalize unfavorable spatial configurations, we introduced a new subgraph of G_{grid}, called the *penalty grid.* This subset of the grid represents the set of points where, if a carbon atom of the benzenic fragment is positioned, a negative energy contribution (i.e., a positive term in the Hamiltonian) is assigned, simulating the effect of an undesired rotation or poor overlap between aromatic planes.

The modified Hamiltonian takes the following form:

$$H_\pi = \frac{\hat{E}}{6} \sum_{i,i'} (\hat{c}_i\, \hat{c}_{i'})\, x_{ii'} \left(\sum_{j,j'} (c_{j'i'} - P\, \tilde{c}_{j'i'})\, c_{ji}\, x_{jj'} \right), \tag{4.10}$$

where $\tilde{c}_{j'i'}$ is a new label assigned to the vertices of G_{grid} [4.7], equal to 1 if node j' belongs to the penalty grid associated with the center $i' \in G_{\text{center}}$ (Figure 4.5), and 0 otherwise. The parameter P represents a penalty coefficient.

Determining the optimal value of P is not straightforward: ideally, it should cancel the positive energy contributions from configurations that are formally allowed within the subgraph G_{benzene}, but correspond to spatial arrangements unfavorable for actual π-stacking interaction. In other words, P should be chosen to neutralize false positive energy contributions due to spatial placements not consistent with a truly significant interaction.

To clarify how the model works, we provide a practical example. Suppose that a benzenic fragment of the ligand is positioned with its center inside the subgraph G_{center}, two of its carbon atoms fall into G_{benzene}, and the remaining four are instead located in the penalty grid G_{penalty}. In this case, the Hamiltonian contribution [4.9] becomes:

$$H_\pi = \frac{\hat{E}}{6}\left(2 - 4P\right).$$

To ensure that such a configuration - considered unfavorable for π-stacking - does not contribute to the total energy, we require $H_\pi = 0$. This implies that the optimal value for the penalty parameter should be $P = \frac{1}{2}$.

However, this approach has some limitations. For example, if three carbon atoms fall into G_{benzene} and three into G_{penalty}, the Hamiltonian would be:

$$H_\pi = \frac{\hat{E}}{6}\left(3 - 3P\right) = \frac{\hat{E}}{6} \cdot \frac{3}{2} = \frac{\hat{E}}{4},$$

which corresponds to a positive energy contribution of unjustified intensity. This result highlights a structural weakness of the proposed model, as it assigns energy to geometrically incompatible configurations that we would prefer to neglect.

Another issue arises when the benzenic fragment of the ligand is positioned orthogonally to the aromatic fragment of the protein, while its center remains inside G_{center}. Although such an orientation does not favor any π-stacking interaction in reality, the six carbon atoms may still fall within the penalty grid G_{penalty}, thus generating a “penalising” energy contribution.

In light of these observations, we deemed it appropriate to abandon this model, as it does not adequately capture the geometric dependence of the π-stacking interaction.

4.4 Third Modeling Approach

We therefore opted for a model that allows us to constrain the benzenic fragment of the ligand exclusively to spatial configurations where it is nearly parallel to the benzenic ring of the protein. To impose this constraint, we adopted a modeling strategy that fixes the normal vector to the plane containing the ligand's benzenic fragment, thus giving up the information related to possible in-plane rotations. However, as already pointed out earlier, the energy variation due to rotation within the fragment's plane is negligible compared to the total interaction energy. Therefore, this approximation is consistent with the assumptions of the overall model and does not significantly compromise the accuracy of the computed π-stacking Hamiltonian.

The adopted strategy consists of adding, for each benzenic fragment of the ligand, a new vertex to the molecular graph G_{mol}. In particular, for each aromatic ring, we introduce an additional node such that the segment connecting it to the center of the fragment is perpendicular to the plane of the ring itself, as shown in Figure 4.7 This mechanism allows us to geometrically represent the direction of the normal vector to the benzenic fragment's plane, enabling us to constrain the spatial orientation of the molecule.

As before, we modify the node labels in the molecular graph G_{mol}, associating the following variables to each node i:

- a variable $\hat{c}_i$, which equals 1 if node i represents the center of a benzenic fragment, and 0 otherwise;
- a variable c_{ij}, which equals 1 if node i is the additional vertex that, together with the center j, defines a segment orthogonal to the fragment's plane, and 0 otherwise.

The enriched structure of the molecular graph thus becomes:

$$G_{\text{mol}} = \{i,\, e_{ij},\, w_{ij},\, q_i,\, \hat{c}_i,\, c_{ij}\}\,. \tag{4.11}$$

As already introduced in the previous modeling approach, we consider a subgraph of G_{grid}, denoted G_{center}, representing the set of grid vertices where

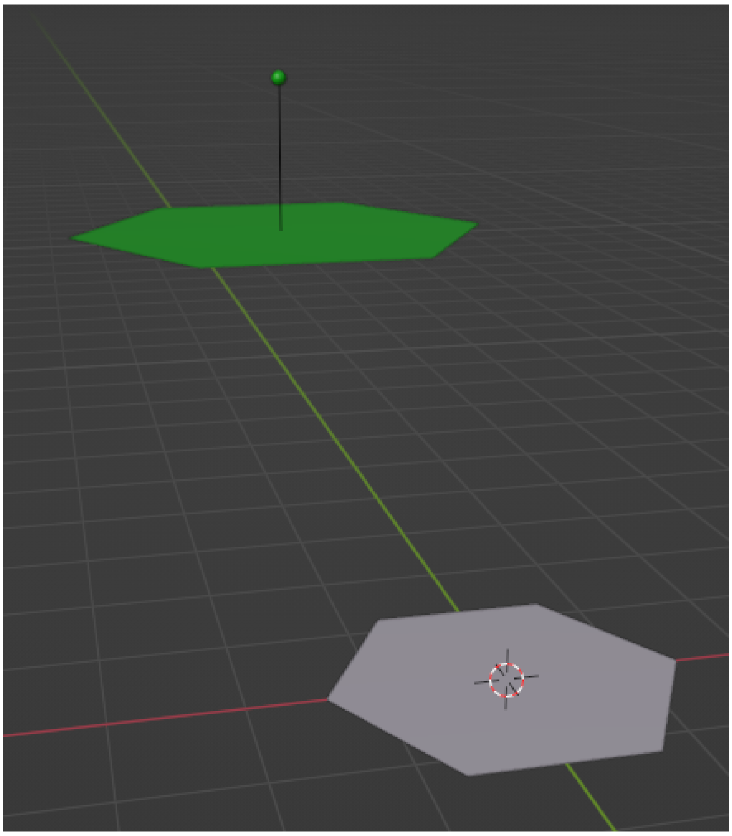


Figure 4.7: Additional node placed along the normal vector to the aromatic plane

the center of a benzenic fragment can be placed to give rise to a favorable π-stacking interaction, provided the fragment's orientation is also compatible.

For each vertex $i' \in G_{\text{center}}$, we define a subset of vertices

$$\{j'_1, j'_2, \ldots, j'_n\} \subseteq G_{\text{grid}},$$

such that the segment $i'j'_k$ forms an angle with the plane of the protein's benzenic fragment lying in the interval $\left[\frac{\pi}{2} - \alpha,\ \frac{\pi}{2} + \alpha\right]$, where $\alpha > 0$ is a tunable parameter that controls the tolerance with respect to perfect orthogonality.

In Figure 4.8, we show a benzenic fragment of the ligand (in green) with its center located within G_{center} and the additional vertex placed inside a sphere representing the set of nodes $\{j'_1, \ldots, j'_n\}$. We observe that, for small rotations of the benzene molecule, the new vertex remains inside the sphere.

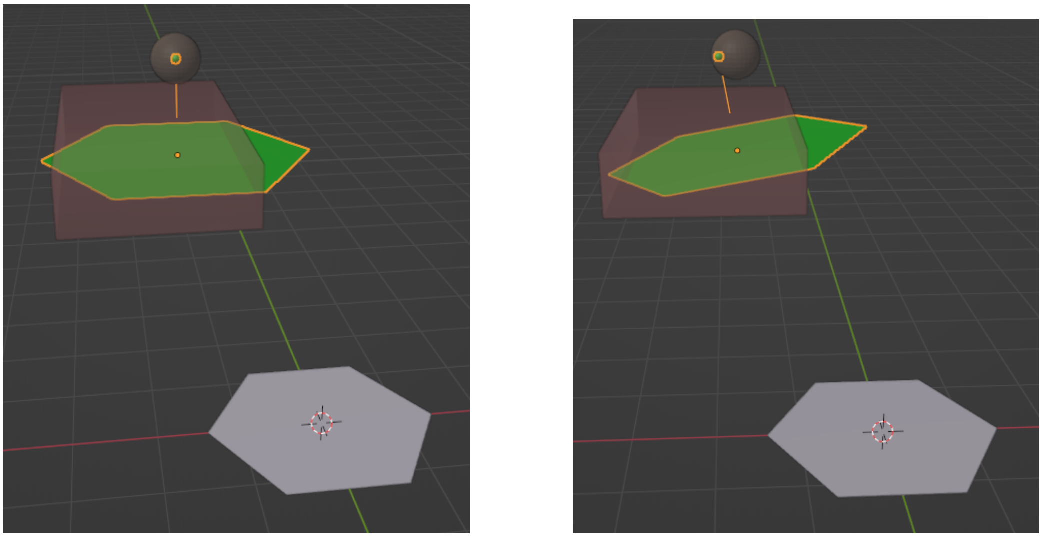


Figure 4.8: Illustration of a benzene fragment from the ligand (in green) with its center placed in G_{center} and an auxiliary node positioned within the relative sphere.

Regarding the grid G_{grid}, we introduce the following labeling system:

- A variable $\hat{c}_{i'}$, which takes the value 1 if vertex i' belongs to the subgraph G_{center}, i.e., if it can represent a possible center for positioning a benzenic fragment of the ligand. Otherwise, $\hat{c}_{i'} = 0$.

- A variable $c_{j'i'}$, which represents the geometric interaction intensity associated with the pair of vertices j' and i', where i' belongs to the subgraph G_{center}. Thanks to the numerical simulations performed (Sec. 3.3.2), we are able to estimate the interaction energy for each relative configuration between two benzene molecules. In particular, by fixing the position of the benzene molecule in the protein and assigning the center of the ligand's benzenic fragment to a point i', we can precisely calculate how the interaction energy varies with the orientation of the ligand's fragment - that is, as a function of its spatial rotations. Once the direction of the normal to the ligand's benzenic fragment is fixed - corresponding to the direction of the oriented segment $i' \to j'$ - we obtain an energy interval associated with possible in-plane rotations of the aromatic ring. By computing the average of this energy interval, we can assign a single scalar value to the pair (j', i') that reflects the average interaction intensity in that configuration. The variable $c_{j'i'}$ is therefore defined as the magnitude of this average energy value, and expresses how effectively the direction $j'i'$ supports a π-stacking interaction. In this way, the label $c_{j'i'}$ compactly and continuously encodes the energetic information obtained from simulations, providing a quantitative basis for assessing the spatial orientation of the ligand's benzenic fragment relative to the protein.

If j' does not belong to the previously introduced set $\{j'_1, \ldots, j'_n\}$ relative to i', we set $c_{j'i'} = 0$, effectively excluding the possibility of a favorable interaction in that direction.

This modeling allows for a more nuanced representation of the spatial alignment quality of the ligand's benzenic fragment with respect to the protein, overcoming the rigidity of a purely binary (acceptable/unacceptable) approach. Thus, the model can assign an energy contribution proportional to the degree of geometric alignment, improving sensitivity in capturing intermediate configurations.

The structure of the grid graph then becomes:

$$G_{\text{grid}} = \{i',\, e_{i'j'},\, w_{i'j'},\, V_{i'},\, \hat{c}_{i'},\, c_{i'j'}\}\,. \tag{4.12}$$

We can now define the new Hamiltonian:

$$H_{\pi} = \sum_{i,i'} \hat{c}_i\, \hat{c}_{i'}\, x_{ii'} \left(\sum_{j,j'} c_{ij}\, c_{i'j'}\, x_{jj'} \right). \tag{4.13}$$

Regarding the *T-shaped* configuration, the simulations shown in Figure 3.3.4 clearly demonstrate that the interaction energy remains nearly constant as the rotation angle of the vertical molecule around the z-axis varies. This behavior suggested a simplification of the model by fixing this axis as

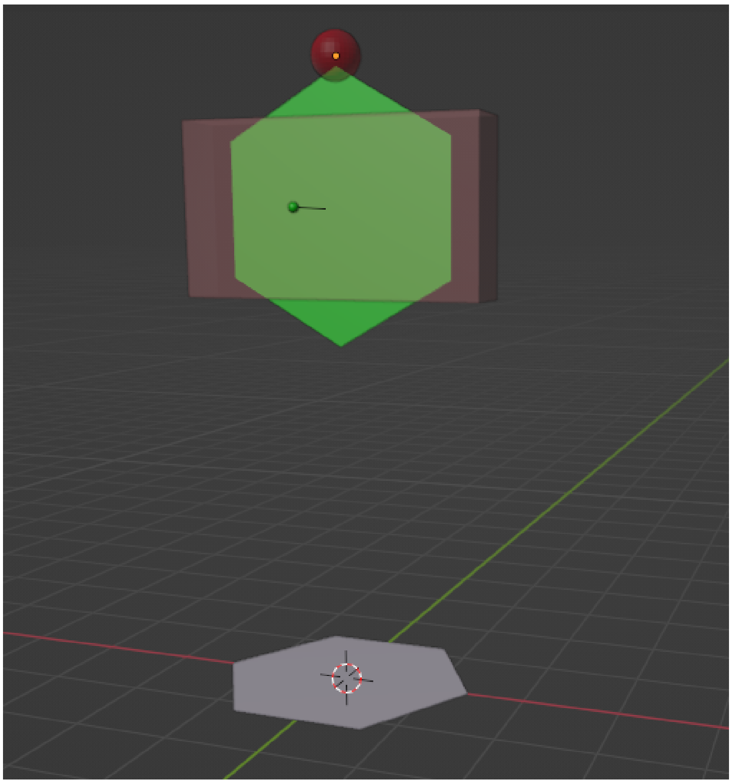


Figure 4.9: Visualization of the T-shaped configuration with fixed vertical axis

the preferred direction for the interaction, similarly to what was already done for the *parallel-displaced* configuration.

We therefore proceed by introducing, for each vertex $i' \in G_{\text{center}}$ associated with the *T-shaped* configuration, a set of nodes

$$\{j'_1, j'_2, \ldots, j'_n\} \subseteq G_{\text{grid}},$$

such that, if the center of the ligand's benzenic fragment is positioned at i' and one of its carbon atoms lies at a node j'_k (as represented in Figure 4.8), then a favorable condition for π-stacking interaction occurs.

Based on this construction, we define the following Hamiltonian associated with the *T-shaped* configuration:

$$H_{\pi,\text{T-shaped}} = \sum_{i,i'} \hat{c}_i \, \hat{c}_{i'} \, x_{ii'} \left(\sum_{j,j'} c_{ij} \, c_{i'j'} \, x_{jj'} \right),$$

where the involved labels have the following meanings:

- $\hat{c}_i = 1$ if node i is the center of a benzenic fragment in the ligand, and 0 otherwise;
- $\hat{c}_{i'} = 1$ if node i' belongs to the subset $G_{\text{centri T-shaped}}$;
- $c_{ij} = 1$ if node j represents a carbon atom of the benzenic fragment centered at i;
- $c_{i'j'}$ is a continuous variable representing the geometric interaction intensity between node j' and center i', as a function of orientation and distance with respect to the target molecule.

This formulation allows us to coherently model the π-stacking interaction in the T-shaped configuration, constraining geometries to physically meaningful configurations according to the evidence provided by the simulations.

Chapter 5

Results

In this chapter, we present the results obtained from the simulated annealing computations carried out using different test molecules.

For the simulated annealing computations presented in this work, we employed the `neal.SimulatedAnnealingSampler` from the `neal` Python package 1.1.3.

5.1 QUBO Coefficient

We now proceed to perform both simulated and quantum annealing calculations, as previously introduced. The central objective is to compute the molecular arrangement that yields the minimum energy for the Hamiltonian introduced in Section 2.1.3, which is enhanced with an additional term that aims to capture the effects of π-stacking.

We chose to include in this thesis only the results related to the parallel-displaced configuration, as those corresponding to the T-shaped configuration are entirely analogous.

It is important to note that the Hamiltonian described in Section 4.4 was parametrized based on the interaction energies obtained from Q-Chem simulations described in Section 3.2. These simulations provide the total interaction energy between the two benzene fragments, which inherently includes all pairwise atomic interactions, as well as any collective contributions arising from molecular effects.

For the purposes of the annealing calculations, we are specifically interested in isolating the portion of the interaction energy that cannot be directly attributed to individual atom-atom interactions. This is necessary because the original Hamiltonian from Section 2.1.3 already explicitly includes atomistic interaction terms, such as Coulomb and van der Waals contributions, in closed-form expressions. Including these contributions again from the Q-Chem-derived π-stacking term would result in double counting.

To address this, we subtract from the total interaction energy obtained from the Q-Chem simulations the analytically known atomistic contributions, namely the Coulomb and van der Waals terms. The resulting quantity, which we refer to as the *latent energy*, or *molecular energy*, represents the residual component of the interaction energy that cannot be explained by simple pairwise atomic forces. This latent energy is assumed to capture the molecular-level effects, including those arising from π-stacking, that are not explicitly accounted for in the original Hamiltonian.

It is also important to clarify that we do not subtract the contributions associated with other known contributions, such as hydrogen bonding or hydrophobic interactions. This is due to that fact that, in two benzene rigs, these contributions do not occur. In particular, hydrogen bonding does not occur, as carbon atoms lack sufficient electronegativity to participate in such interactions. Similarly, hydrophobic interactions are a solvent-mediated phenomenon that occurs in aqueous environments, where nonpolar molecules tend to aggregate to minimize their contact with water. Since the Q-Chem simulations were conducted in vacuum conditions, removing a hydrophobic term would be inappropriate, as this interaction is not present in the simulation data to begin with.

This correction procedure ensures that the π-stacking term incorporated into the annealing Hamiltonian reflects only the non-atomistic, collective contributions relevant to the molecular system under investigation.

To quantify the Coulomb and van der Waals components, we developed a computational code capable of calculating all individual atom-atom interaction terms for each spatial configuration and orientation of the two benzene molecules, using the parameters derived from MMFF94 [40]. By summing these pairwise contributions, the code computes the total electrostatic plus van der Waals energy associated with each configuration. This calculated energy is then subtracted from the total interaction energy obtained from the Q-Chem simulations, isolating the residual contribution relevant for our modeling purposes.

5.1.1 Calculation of coefficients and results

We chose to perform these calculations only for configurations where the total interaction energy from the simulations was lower than $-2\,\mathrm{kcal/mol}$. This threshold was selected to ensure that only configurations exhibiting a significant and physically meaningful π-stacking interaction were considered. As shown in Figure 3.3, this threshold effectively excludes the local maxima of each energy–displacement curve at constant vertical separation. In fact, all curves exhibit a peak at zero lateral displacement, and a less stringent cut-off (e.g. $-1\,\mathrm{kcal/mol}$) would cause some of these peaks to be retained in the dataset. We therefore discard such contributions in order to focus exclusively on those configurations that are genuinely favorable for each fixed interplanar distance. Moreover, including higher-energy configurations, where the interaction is weak or negligible, could introduce noise and obscure the energetic signature of π-stacking , as stated in section 3.3.1 .

It is important to note that the simulation results display a continuous energy trend with respect to molecular displacement. As the system deviates from its equilibrium configuration, the interaction energy gradually decreases in magnitude. Therefore, setting an energy threshold was necessary to discriminate between configurations where the π-stacking interaction dominates.

We chose to compute the electrostatic interaction component and van

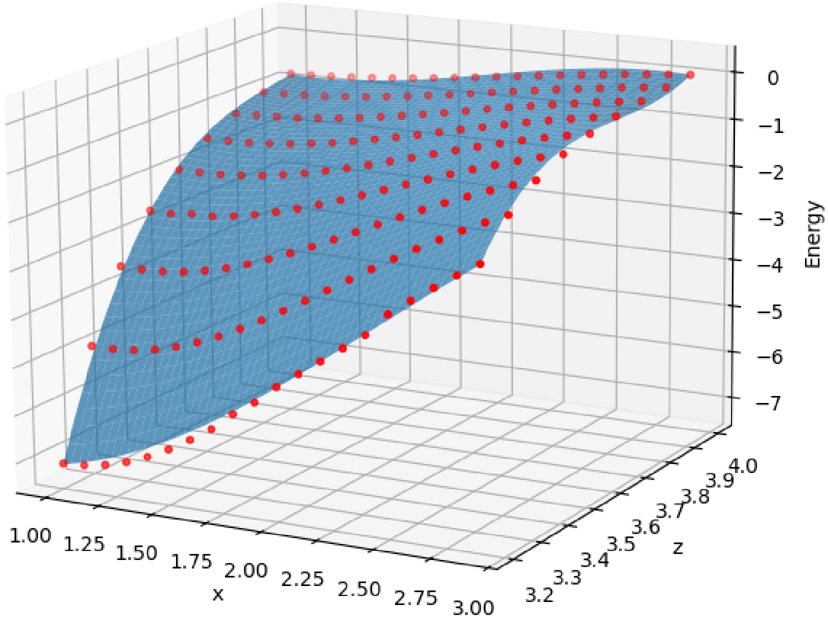


Figure 5.1: Variation of *molecular energy* with respect to lateral displacement and vertical distance

der Waals fluctuations by performing calculations for each individual configuration previously sampled in our simulations. Subsequently, we subtracted these computed energetic contributions from the total interaction energies obtained via simulation. This procedure yielded a set of values representing the *molecular energy*, which we used to determine the coefficients of our QUBO formulation. In Figure 5.1, we provide a graphical representation of the intensity of the resulting residual molecular energy.

The data points obtained were then fitted using a third-degree polynomial in the variables d (displacement) and z (interplanar distance). The resulting polynomial model is given by:

$$\begin{aligned} f(d,z) = & -0.413\,d^3 + 12.322\,z^3 - 0.730\,d^2z + 4.638\,dz^2 + 5.341\,d^2 \\ & -150.773\,z^2 - 33.423\,dz + 55.572\,d + 613.092\,z - 827.530. \end{aligned}$$

The goodness of fit is confirmed by the coefficient of determination, which is $R^2 = 0.998$.

Since the modeling approach presented in 4.4 does not account for rotations of the two molecules around the z-axis in the construction of the Hamiltonian, we chose to rely exclusively on the interaction energy as a function of the lateral displacement and the vertical distance between the molecular planes. In particular, we decided not to use the energy expressed as a function of the planar radial distance, because modifying the y-coordinate of one of the two molecules effectively corresponds to a combination of increasing the interplanar distance and applying a rotation around the z-axis. Such a transformation does not alter the intrinsic interaction in a way that is consistent with the assumptions of our Hamiltonian, and would introduce rotational effects that are deliberately excluded from our current model. Therefore, to preserve coherence with the theoretical framework adopted, we restrict our analysis to configurations parameterized only by vertical and lateral displacements within the plane.

Due to the high computational cost and the time required to obtain reliable results, we chose to compute the interaction energy associated with rotational configurations only at the equilibrium position. From the corresponding simulations, we estimated the percentage variation in energy attributable to molecular rotations. Assuming that this variation remains approximately constant across all spatial configurations considered, we extrapolated the interaction energy for each configuration by scaling the energy obtained at the molecular center by the previously determined rotational correction factor. This approximation allowed us to efficiently incorporate the effect of rotations into the model while significantly reducing the number of simulations needed. In any case, in the next stage of research, we reserve the possibility of conducting more extensive simulations that would allow us to validate our current assumption and, if necessary, obtain more accurate and reliable results.

In Figure 5.2 the percentage variations in interaction energy associated with rotations around the x-axis -corresponding to the direction of lateral displacement - denoted by the angle α, and around the y-axis, denoted by the angle β. Due to the symmetry of the molecular configuration with respect

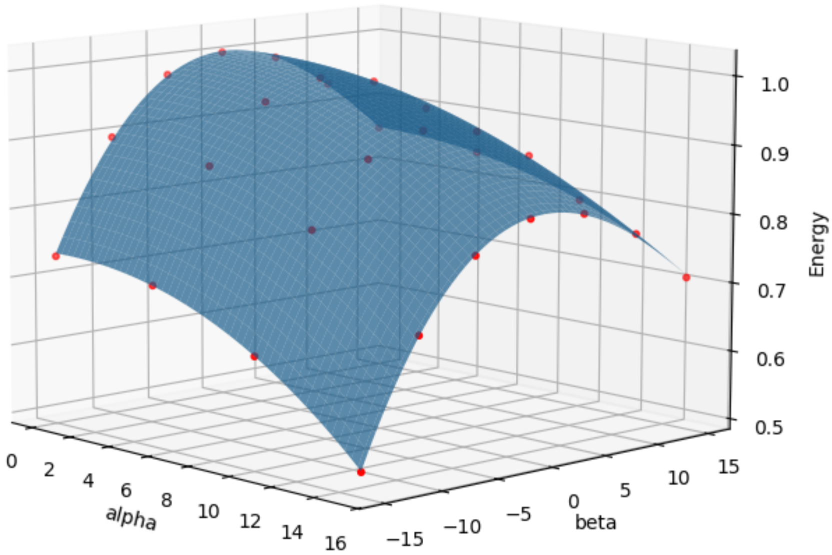


Figure 5.2: Percentage variations in interaction energy associated with rotations around the x and y-axis

to positive and negative rotations around the displacement axis, the values of α are taken to be strictly positive. In contrast, the angle β also includes negative values, as no such symmetry constraint applies to rotations around the y-axis.

The following polynomial equation interpolates the latent energy data presented above: The following polynomial equation interpolates the latent energy data presented above:

$$\begin{aligned} g(\alpha,\beta) = & -8.95\times10^{-6}\,\alpha^3 + 1.57\times10^{-5}\,\beta^3 + 1.34\times10^{-5}\,\alpha^2\beta - 2.54\times10^{-6}\,\alpha\beta^2 \\ & -6.71\times10^{-4}\,\alpha^2 - 9.15\times10^{-4}\,\beta^2 - 4.79\times10^{-5}\,\alpha\beta \\ & +6.64\times10^{-4}\,\alpha + 5.15\times10^{-4}\,\beta + 1.00. \end{aligned}$$

The coefficient of determination is $R^2 = 0.999$.

In conclusion, to determine the coefficients $c_{i'j'}$ in accordance with the notation introduced in Section 4.13, we proceed through the following steps:

- We identify all benzene fragments sufficiently close to the binding pocket and compute their centroids C along with the normal vectors to the corresponding aromatic planes.

- For each such fragment, we search for all grid points P whose vertical distance z from the aromatic plane lies between 3.2 Å and 4 Å, and whose planar radial displacement d lies between 1 Å and 3 Å. The computation of these distances is carried out relative to the plane of the protein's aromatic fragment and its associated normal vector.

- Once these points P are identified, we check for the existence of suitable *anchor points* in the vicinity, denoted by V. These anchor points represent the additional vertical nodes required to define the rotation of the aromatic fragment. We consider as valid anchors those located within a distance of 1–2 Å from the corresponding point P, and for which the segment connecting P and V forms an angle of at most 15 degrees with respect to both the displacement axis and its orthogonal direction, when projected onto the plane orthogonal to the normal vector of the protein fragment. These angles are denoted by α and β, respectively.

- Finally, we assign the coefficient $c_{i'j'}$ as the product $f(d, z) \cdot g(\alpha, \beta)$, where f and g are those previously defined.

5.2 Simulated Annealing Calculations

To test the proposed model, we chose to perform calculations on small *test* molecules, allowing for a direct and immediate verification of the results. Indeed, when evaluating the model, it is common to use protein pockets - fragments of the protein that surround the binding site and typically contain hundreds of atoms. Within such environments, identifying the dominant interaction mechanisms becomes particularly challenging due to the complexity and density of atomic contributions.

A central role in addressing this issue is played by the multiplicative parameters λ, introduced in Section 2.1.3, whose purpose is to appropriately rescale the different energetic contributions in order to make them comparable and meaningful. For instance, the geometric contribution in the Hamiltonian typically attains very large values (on the order of 50), especially when the grid points do not precisely match the atomic positions of the ligand. In contrast, the contribution introduced in this work - modeling the π-stacking interaction - takes values ranging approximately from -7 to 0. Without the introduction of a suitable λ-parameter, this term would remain largely negligible when compared to other dominant components of the energy function. Therefore, identifying the optimal values of λ is essential to correctly account for the π-stacking contribution within the QUBO formulation.

In the experimental phases following the publication of this thesis, we intend to focus on calibrating the λ parameters to properly balance and integrate the energetic terms proposed herein.

5.2.1 Ligand, and pockets

To validate our model, we first examined the interaction between two benzene rings, which served as a reference system. We then extended our analysis to the interaction between a simple benzene fragment and three aromatic amino acids: tyrosine, phenylalanine, and tryptophan. These amino acids, which explicitly feature aromatic rings, are among the most frequently reported in the literature for their involvement in π-π interactions. Several studies have highlighted the role of aromatic amino acids in non-covalent interactions. In particular, Shao et al. [41] and Burley and Petsko [42] have shown that tyrosine, tryptophan, and phenylalanine frequently participate in such interactions.

At first, we added the aromatic center and the vertical anchoring node to the ligand structure. Naturally, these nodes - being *fictitious*, i.e., not corresponding to actual atoms - do not carry any physical contributions such as partial charges or van der Waals parameters. The vertical node was placed at a distance of 1.5 Å from the center of the benzene ring.

We present a graphical representation of the ligand in Figure 5.3

For what regards amino acids, tyrosine is an amino acid that plays a

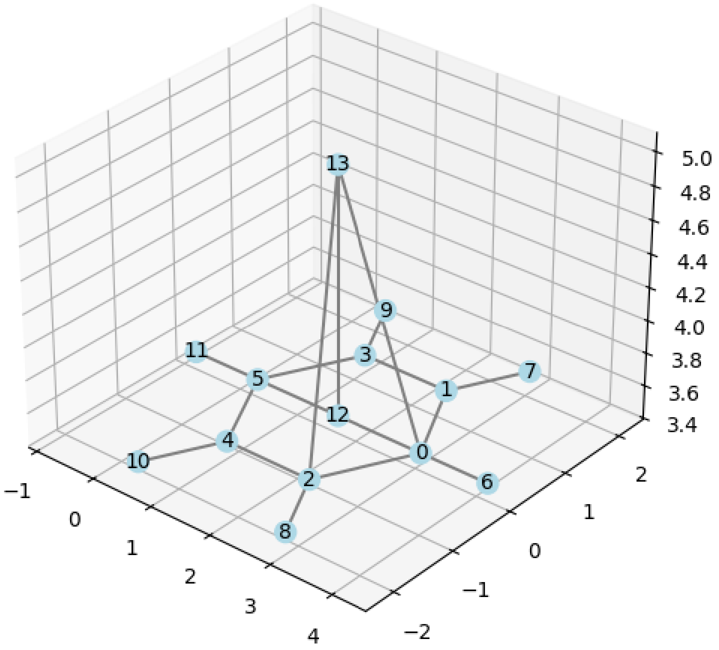


Figure 5.3: Benzene ligand with auxiliary nodes

crucial role in the human body, acting as a precursor to important molecules such as neurotransmitters, melanin, and thyroid hormones. Its molecular formula is $C_9H_{11}NO_3$, and its three-dimensional structure is shown in figure 5.4 (a).

Tyrosine contains a phenol functional group, consisting of a benzene ring substituted with a hydroxyl group (OH) in the para position relative to the side chain. Note that in aromatic chemistry, the *para* position refers to the 1,4-substitution pattern on a benzene ring. That is, when two substituents are located on carbon atoms separated by two positions (e.g., positions 1 and 4), they are considered to be in *para* configuration.

To further validate the proposed model, we repeated the calculations by replacing the tyrosine fragment with a phenylalanine fragment $C_9H_{11}NO_2$ (Figure 5.4 (b)).

Phenylalanine features a phenyl group (C_6H_5) directly attached to the side chain. Unlike tyrosine, it lacks substituents on the benzene ring. It is a purely hydrophobic aromatic system capable of engaging in π-stacking interactions. Lastly, we also carried out the calculations for the tryptophan–benzene interaction. Tryptophan ($C_{11}H_{12}N_2O_2$, Figure 5.4 (c)) is an essential amino acid and it is a fundamental component of proteins and plays a crucial role in the synthesis of important biological compounds, including serotonin and melatonin. Tryptophan contains an indole group, which consists of a benzene ring fused to a five-membered nitrogen-containing pyrrole ring. The aromatic benzene fragment is thus embedded within a larger conjugated system, making it chemically distinct from an isolated phenyl ring.

All three amino acids considered in this study contain an aromatic moiety, that is, a structurally distinct portion of the molecule characterized by a conjugated ring system with delocalized π electrons.

It is important to note that amino acids do not contain isolated benzene molecules; rather, they possess aromatic fragments integrated into their molecular structure. These aromatic moieties, such as phenyl groups in phenylalanine, phenol groups in tyrosine, or indole systems in tryptophan, are responsible for enabling π-stacking interactions. Although structurally embedded within larger frameworks, these fragments retain the characteristic electronic properties of aromatic systems and significantly contribute to non-covalent interactions relevant in biomolecular contexts.

5.2.2 Pocketgrids

To rigorously evaluate our model, we designed a synthetic pocketgrid centered around the pocket molecule.

Initially, we constructed a minimal grid configuration including only the specific nodes that we intended to be occupied by the ligand. This allowed us to work with a highly simplified and controlled environment, suitable for verifying the correctness of our modeling assumptions. We will refer to this setup as *Grid A*.

More precisely, *Grid A* consists of 14 nodes: 6 nodes correspond to the positions of the carbon atoms in the aromatic ring, 6 nodes to the associated

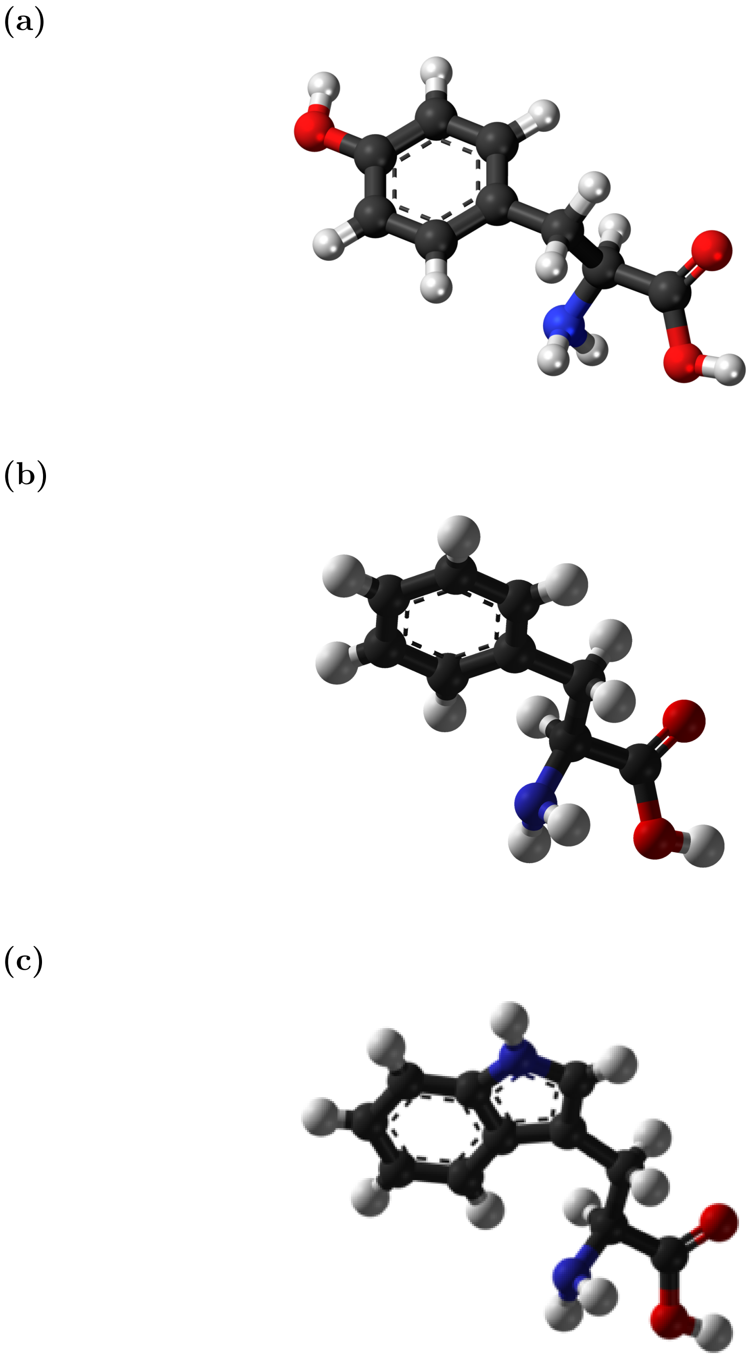


Figure 5.4: (a) Tyrosine (b) Phenylalanine (c) Tryptophan. Images created using *BIOVIA Discovery Studio Visualizer*.

hydrogen atoms, 1 node represents the center of the aromatic fragment, and 1 node is placed vertically above the center to serve as an anchoring point for controlling ring orientation. The center node was positioned at a radial distance of approximately 1.7 Å from the aromatic plane and a vertical offset of about 3.4 Å, with the segment connecting the center to the vertical node aligned parallel to the normal vector of the molecule aromatic fragment.

Subsequently, in order to relax the spatial constraints imposed on the ligand and simulate a more realistic docking scenario, we enriched the pocket by adding several additional nodes in its vicinity. In particular, we randomly inserted 26 extra nodes, including a replica of the original 14-node ligand-compatible arrangement described above, effectively introducing a second ideal candidate position for the aromatic fragment. We will refer to this extended configuration as *Grid B*.

To further expand the search space and assess the robustness of our model under more complex conditions, we then introduced a third set of 14 nodes, also structured to represent another geometrically favorable position for the benzene fragment. This step was instrumental in testing the model's ability to consistently identify optimal binding poses even in the presence of multiple plausible interaction sites. We will refer to this configuration as *Grid C*.

5.2.3 Calculations and results

Regarding the energetic contributions considered, we included the geometric term, the electrostatic interaction, the van der Waals term, and, of course, the molecular term, mainly accounting for π-stacking interaction. In this initial test case, we chose to neglect both the hydrogen bonding and hydrophobic contributions.

Benzene - Benzene

Initially, we chose to consider only a benzene fragment as pocket, neglecting all other atoms.

Table 5.1 summarizes the results obtained for a test configuration where both the ligand and the molecule are benzene, and the pocketgrid used is *Grid A* - a minimal configuration designed to tightly fit the ligand.

The simulation was performed using a set of `Lambda Weights` equal to $[1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 1.0]$, where each value controls the relative contribution of one of the physical or geometric energy terms in the global Hamiltonian. The annealing process was carried out with 100 reads and 2000 sweeps, which proved sufficient for achieving convergence in this simplified environment.

In simulated annealing the term “sweep” denotes the fixed number of move-attempts performed at a single temperature before that temperature is reduced. Each sweep typically consists of a sequence of proposals - one per degree of freedom or a constant multiple thereof - so as to allow the system to equilibrate locally and to overcome energy barriers characteristic of the current thermal stage. By contrast, a “read” designates an entire annealing trajectory, from the initial high temperature down to the final low temperature, executed from a given random or predefined starting configuration. Multiple reads, each with its own sequence of temperature steps and associated sweeps, are carried out in order to sample different regions of the solution space and to reduce the risk of entrapment in suboptimal local minima. The collection of final states obtained from all reads is then examined and the configuration with the lowest recorded energy is selected as the approximate global optimum.

The problem size includes 196 non-zero linear terms and 11,101 non-zero quadratic couplings, for a total of 11,297 active terms in the QUBO formulation. The selected couplers highlight the optimal pairwise node assignments between the ligand and the pocketgrid, indicating a correct and coherent mapping of the benzene atoms onto the grid.

Immediately below the line *“Total non zero terms"* we can observe the ligand-to-grid node matchings corresponding to the optimal configuration. The ligand node indices follow the same numbering as figure 5.3, while the grid nodes are indexed such that nodes 0–5 represent the positions for the carbon atoms, nodes 6–7 correspond to the aromatic center and the vertical anchor, and nodes 8–13 to the hydrogen atoms.

From an energetic perspective, the system achieved a total energy of approximately -3.39 kcal/mol. The breakdown of the contributions is as follows: the geometric energy is near zero (~ 0.003), indicating excellent structural compatibility with the ideal geometry; the van der Waals term contributes -2.31, and the molecular term, mainly accounting for π-stacking interactions, contributes significantly with -2.66. The electrostatic energy, while slightly positive ($+1.58$), does not overcome the stabilizing effects of the other components.

This result confirms that, in the controlled setting of Grid A, the model correctly identifies the intended pose of the ligand, with the molecular term playing a central role in stabilizing the configuration.

From the results shown in Table 5.1, we observe that the geometric energy is nearly zero - as expected - and the sum of the molecular term, van der Waals, and electrostatic contributions is close to $-3\,\mathrm{kcal/mol}$, which matches the energy predicted by the simulations shown in Section 3.2.

Then, we repeated the calculation using *Grid B*, increasing the number of reads and sweeps to 500 and 10,000, respectively, in order to identify the optimal configuration. As the size of the pocket grid increases, the simulated annealing algorithm requires more time to converge to the global minimum of the Hamiltonian. In Table 5.2, we observe that the optimal position of the ligand remains essentially unchanged compared to the previous case, with the only noticeable difference being a rotation of the ligand around the axis defined by the aromatic center and the vertical node - an axis for which we explicitly chose not to apply any rotational penalty. Importantly, the energetic contributions remain virtually unaltered, confirming the robustness of the interaction under the added degrees of spatial freedom.

Grid C causes the simulated annealing process to struggle in identifying the global minimum when the number of reads and sweeps is kept constant.

Table 5.1: Results of the Simulated Annealing computation performed using a benzene ligand, a benzene molecule, and Pocketgrid A.

Configuration	
Lambda Weights	[1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 1.0]
Number of Reads	100
Number of Sweeps	2000
Problem Size	
Non-zero linear terms	196
Non-zero quadratic terms	11 101
Total non-zero terms	11 297
Selected couplers	$\{(9,13),(0,3),(10,9),(11,8),(12,6),(13,7),(1,4)$ $(2,2),(7,12),(3,5),(4,1),(5,6),(6,11),(8,10)\}$
Energies	
Best Energy	-3.3938571373910733
Geometric Energy	0.003140345434058167
vdW Energy	-2.314786993062496
Electrostatic Energy	1.5819669974230242
Molecular Energy	-2.664177487185666

In particular, while the minimum-energy configuration itself does not change with the increased number of nodes, the algorithm requires a greater number of steps to successfully locate the global minimum. This highlights the growing complexity of the energy landscape as the dimensionality of the system increases, and suggests the need for tuning annealing parameters accordingly. In Table 5.3 we observe that the simulated annealing algorithm fails to identify the optimal ligand placement by incorrectly mapping ligand node 10 to grid node 44 and node 7 to node 37.

Nonetheless, it correctly fixes the aromatic center and the vertical anchoring node in the expected positions, thereby preserving the favorable contribu-

Table 5.2: Results of the Simulated Annealing computation performed using a benzene ligand, a benzene molecule, and Pocketgrid B.

Configuration	
Lambda Weights	[1.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	574
Non-zero quadratic terms	92 289
Total non-zero terms	92 863
Selected couplers	$\{(0,1),(5,4),(10,11),(11,12),(4,3),(13,7),(1,0)\}$ $\{(12,6),(2,2),(6,9),(7,8),(3,5),(8,10),(9,13)\}$
Energies	
Best Energy	-3.388727580462687
Geometric Energy	0.008269724178869398
vdW Energy	-2.314786993062496
Electrostatic Energy	1.5819669974230242
Molecular Energy	-2.664177487185666

tion from the derived from the molecular term. This indicates that, although geometric alignment may be partially compromised, the algorithm still prioritizes key interaction features when guided by dominant energetic terms. In order to preserve the ligand geometry more strictly, it would be sufficient to increase the corresponding multiplicative parameter λ. This would enhance the weight of the geometric term in the Hamiltonian, effectively penalizing any deviations from the ideal spatial arrangement and guiding the optimization process toward configurations that maintain the structural integrity of the ligand. Indeed, setting the multiplicative coefficient of the geometric Hamiltonian equal to 3 is sufficient to obtain a mapping analogous to the

Table 5.3: Results of the Simulated Annealing computation performed using a benzene ligand, a benzene molecule, and Pocketgrid C.

Configuration	
Lambda Weights	[1.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 382
Total non-zero terms	166 152
Selected couplers	$\{(0,4),(2,3),(11,9),(10,44),(4,2),(1,5),(3,0)\}$ $\{(12,6),(13,7),(8,11),(5,1),(6,12),(7,37),(9,8)\}$
Energies	
Best Energy	-1.3872916321860629
Geometric Energy	2.5486193260826497
vdW Energy	-2.4237688636220103
Electrostatic Energy	1.1520353925389477
Molecular Energy	-2.664177487185666

one achieved with *Grid B*. In Figure 5.5, we show the actual placement of the ligand within pocketgrid *Grid C*. The green nodes represent the benzene molecule used to define the pocket, the light blue nodes correspond to the pocketgrid, and the red nodes indicate the final position of the ligand after the simulated annealing procedure.

Tyrosine - Benzene

We now present the results of the simulated annealing using a tyrosine fragment as pocket. At first, we did the calculations using Grid B. In Table

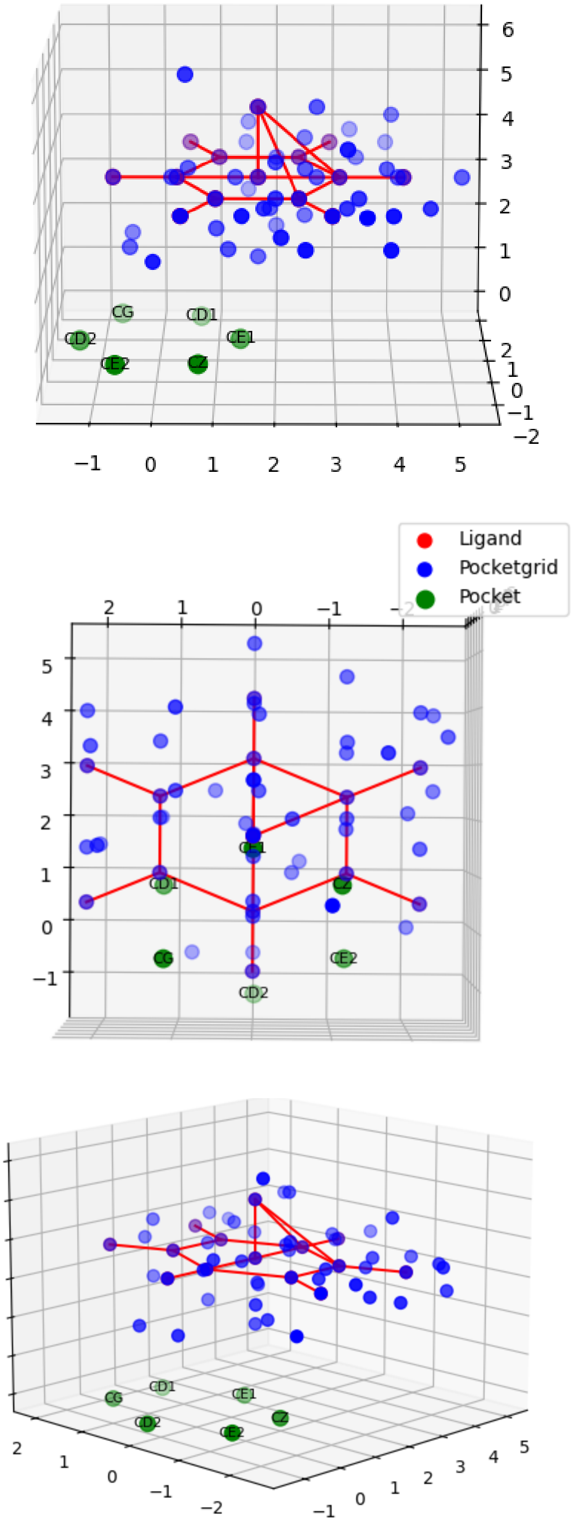


Figure 5.5: Final placement of the ligand within pocketgrid *Grid C* after simulated annealing. Green nodes represent the benzene molecule used to define the pocket, light blue nodes indicate the pocketgrid, and red nodes correspond to the ligand.

Table 5.4: Results of the Simulated Annealing computation performed using a benzene ligand, a tyrosine molecule, and Pocketgrid B.

Configuration	
Lambda Weights	[1.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	574
Non-zero quadratic terms	92 289
Total non-zero terms	92 863
Selected couplers	$\{(11,8),(0,3),(4,1),(1,4),(5,0),(13,7),(12,6),(2,2)$ $(7,12),(3,5),(10,9),(6,11),(8,10),(9,13)\}$
Energies	
Best Energy	-4.632910852359592
Geometric Energy	0.003140345434058167
vdW Energy	-3.057246531658616
Electrostatic Energy	1.08563439979714635
Molecular Energy	-2.664177487185666

5.4, we observe a slight variation in the energy values; however, the ligand positions itself on the grid in such a way as to activate the molecular interaction term.

Moreover, convergence is achieved without the need to modify the annealing parameters, as the addition of atoms to the molecule does not significantly increase the complexity of the optimization process. What primarily influences the computational effort is instead the number of ligand atoms and the number of grid nodes.

After adding additional nodes to the pocket to construct Grid C, we observe - consistent with previous examples - that the simulated annealing

Table 5.5: Results of the Simulated Annealing computation performed using a benzene ligand, a tyrosine molecule, and Pocketgrid C.

Configuration	
Lambda Weights	[1.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 382
Total non-zero terms	166 152
Selected couplers	$\{(0,3),(11,41),(1,4),(12,6),(13,7),(10,9),(2,2)\}$ $\{(3,5),(7,12),(4,1),(5,0),(6,11),(8,10),(9,21)\}$
Energies	
Best Energy	-4.017530851045716
Geometric Energy	-1.2784469779583828
vdW Energy	-3.155647987894283
Electrostatic Energy	0.5238713452739902
Molecular Energy	-2.664177487185666

algorithm has difficulty finding the minimum-energy configuration, as can be seen in Table 5.5.

However, it can be seen in Table 5.6 that by imposing the multiplicative coefficient $\lambda_{\text{geom}} = 3.0$, while keeping the number of reads and sweeps fixed, we are still able to recover the minimum-energy configuration.

In Table 5.7, we present the results obtained by performing the computation while explicitly neglecting the molecular term term - that is, by forcing its contribution to zero. As a result, it becomes immediately evident that the aromatic fragment occupies a different position within the grid. In this case as well, we constrained the ligand geometry by setting the corresponding

Table 5.6: Results of the Simulated Annealing computation performed using a benzene ligand, a tyrosine molecule, and Pocketgrid C (with $\lambda_{\text{geom}} = 3.0$).

Configuration	
Lambda Weights	[3.0, 1.0, 1.0, 0.0, 1.0, 0.0, 1.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 382
Total non-zero terms	166 152
Selected couplers	$\{(3,1),(10,13),(11,8),(0,3),(9,9),(12,6),(13,7)\}$ $\{(1,2),(2,4),(7,10),(4,5),(5,0),(6,11),(8,12)\}$
Energies	
Best Energy	-4.626367382727482
Geometric Energy	-0.092442053302174501
vdW Energy	-3.0572465361585616
Electrostatic Energy	1.0856393979746346
Molecular Energy	-2.664177487185666

multiplicative coefficient to 3, in order to assess the placement of the benzene ring under strict geometric constraints. From this result, we can infer that the molecular term plays a decisive role in determining the optimal positioning of the aromatic fragment with respect to the interacting molecule. As an additional test, we performed the calculations neglecting the molecular term and without enforcing strict geometric constraints - that is, by setting the multiplicative coefficient $\lambda = 1$ for the geometric term. As in previous cases, the simulated annealing algorithm fails to identify the minimum-energy configuration when using the same number of reads and sweeps, as can be seen in Table 5.8. Moreover, even under these relaxed geometric conditions, the

Table 5.7: Results of the Simulated Annealing computation performed using a benzene ligand, a tyrosine molecule, and Pocketgrid C (with $\lambda_{\text{geom}} = 3.0$ and $\lambda_{\text{mol}} = 0.0$).

Configuration	
Lambda Weights	[3.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 379
Total non-zero terms	166 149
Selected couplers	$\{(5,41),(3,46),(0,44),(10,50),(11,49),(13,48),(12,47)\}$ $\{(8,51),(1,45),(7,53),(2,43),(4,42),(6,52),(9,54)\}$
Energies	
Best Energy	-3.127581039752514
Geometric Energy	0.094238755649100843
vdW Energy	-2.785603726443116
Electrostatic Energy	-0.35136755309100084
Molecular Energy	0.0

ligand does not localize within the region expected for favorable π-stacking interactions. This further highlights the critical importance of the molecular term contribution, which proves to be essential even in systems involving small molecules.

Phenylalanine - Benzene

As can be seen in Table 5.9, the phenylalanine fragment also yields results consistent with previous cases, as the molecular term remains the dominant

Table 5.8: Results of the Simulated Annealing computation performed using a benzene ligand, a tyrosine molecule, and Pocketgrid C (with $\lambda_{\text{geom}} = 1.0$ and $\lambda_{\text{mol}} = 0.0$).

Configuration	
Lambda Weights	[1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 379
Total non-zero terms	166 149
Selected couplers	$\{(9,50),(0,46),(10,52),(11,51),(2,45),(12,35),(13,48)\}$ $\{(1,34),(3,42),(4,44),(5,43),(6,54),(7,14),(8,53)\}$
Energies	
Best Energy	-0.9405758111578422
Geometric Energy	3.018710537609026
vdW Energy	-2.6461049468682454
Electrostatic Energy	-1.3131769120839338
Molecular Energy	0.0

factor in determining the spatial configuration of the aromatic moiety.

As with tyrosine, neglecting the molecular term, in the case of phenylalanine results in a different minimum-energy configuration, as presented in Table 5.10.

Tryptophan - Benzene

Also for tryptophan the results clearly highlight the significant role played by the molecular term, as presented in Table 5.11.

Table 5.9: Results of the Simulated Annealing computation performed using a benzene ligand, a phenylalanine molecule, and Pocketgrid C (with $\lambda_{\text{geom}} = 3.0$).

Configuration	
Lambda Weights	[3.0, 1.0, 1.0, 0.0, 0.0, 0.0, 1.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 381
Total non-zero terms	166 151
Selected couplers	$\{(11,8),(0,3),(1,4),(12,6),(13,7),(10,9),(2,2)\}$ $\{(3,5),(7,12),(4,1),(5,0),(9,13),(6,11),(8,10)\}$
Energies	
Best Energy	-4.700995158883253
Geometric Energy	-0.0924410633642175
vdW Energy	-3.657509237261247
Electrostatic Energy	0.6431992581567876
Molecular Energy	-2.296025576075945

When the molecular term is artificially set to zero, the resulting minimum-energy configuration changes once more, as can be seen in Table 5.12, further confirming the essential role of the interaction term introduced in our model.

Table 5.10: Results of the Simulated Annealing computation performed using a benzene ligand, a phenylalanine molecule, and Pocketgrid C (with $\lambda_{\text{geom}} = 3.0$ and $\lambda_{\text{mol}} = 0.0$).

Configuration	
Lambda Weights	[3.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0]
Number of Reads	500
Number of Sweeps	10000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 379
Total non-zero terms	166 149
Selected couplers	$\{(5,41),(3,46),(0,44),(10,50),(11,49),(13,48),(12,47)\}$ $\{(8,51),(1,45),(7,53),(2,43),(4,42),(6,52),(9,54)\}$
Energies	
Best Energy	-2.8897342861621154
Geometric Energy	0.009423875546901828
vdW Energy	-2.74464882884347
Electrostatic Energy	-0.15461126112561147
Molecular Energy	0.0

Table 5.11: Results of the Simulated Annealing computation performed using a benzene ligand, a tryptophan molecule, and Pocketgrid C (with $\lambda_{\text{geom}} = 3.0$).

Configuration	
Lambda Weights	[3.0, 1.0, 1.0, 0.0, 0.0, 0.0, 1.0]
Number of Reads	100
Number of Sweeps	2000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 382
Total non-zero terms	166 152
Selected couplers	$\{(0,0),(2,5),(11,11),(10,12),(3,2),(7,9),(12,6),(13,7)\}$ $\{(1,1),(8,13),(4,4),(5,3),(6,8),(9,10)\}$
Energies	
Best Energy	-4.346754590768584
Geometric Energy	0.0094215480205753365
vdW Energy	-3.275729484362571
Electrostatic Energy	1.31605458889251902
Molecular Energy	-2.3908170354769775

Table 5.12: Results of the Simulated Annealing computation performed using a benzene ligand, a tryptophan molecule, and Pocketgrid C (with $\lambda_{\text{geom}} = 1.0$ and $\lambda_{\text{mol}} = 0.0$).

Configuration	
Lambda Weights	[3.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0]
Number of Reads	100
Number of Sweeps	2000
Problem Size	
Non-zero linear terms	770
Non-zero quadratic terms	165 379
Total non-zero terms	166 149
Selected couplers	$\{(5, 42), (6, 53), (3, 43), (0, 45), (10, 49), (2, 46), (11, 50), (13, 48)\}$ $\{(12, 47), (9, 16), (1, 44), (4, 41), (8, 54), (7, 52)\}$
Energies	
Best Energy	-1.8167994010083414
Geometric Energy	0.4506786219562666
vdW Energy	-2.5807720222992677
Electrostatic Energy	0.313204300343465454
Molecular Energy	0.0

Conclusions

This thesis addressed the mathematical modeling of molecular interactions through a discrete Hamiltonian framework, explicitly written in QUBO form, in which all interaction terms and constraints are expressed using only binary variables and quadratic expressions. The use of this formulation was motivated by its compatibility with quantum annealing algorithms, which are designed to efficiently solve combinatorial optimization problems. In particular, our objective was to apply this approach to molecular docking, where the goal is to determine the most favorable configuration of a small molecule (ligand) within the binding pocket of a target protein. In this context, the lowest-energy configuration is assumed to correspond to the biologically relevant binding pose.

The focus of the work was on modeling non-covalent π-stacking interactions between aromatic rings. The central aim was to define a Hamiltonian that is both mathematically well-posed and physically meaningful, capable of encoding the essential features of π-stacking in a structure suitable for discrete optimization over graph-based representations.

To inform and calibrate the construction of the model, we carried out a series of quantum chemical simulations on representative molecular systems, including benzene dimers and interactions with aromatic amino acids. These simulations provided reference energy landscapes against which the structure and coefficients of the Hamiltonian were tuned. The final model extends the geometric formulation proposed by Triuzzi et al [1] by introducing a new interaction term that captures both the directional and distance-dependent nature of π-stacking interactions.

The resulting QUBO Hamiltonian was tested using classical simulated annealing, due to the temporary unavailability of the D-Wave quantum annealer for technical reasons. Despite this limitation, the methodology proved effective: the simulated annealing algorithm successfully identified low-energy configurations consistent with the reference data, validating the correctness and applicability of the model. Future work will involve repeating the simulations using quantum annealing hardware, with the aim of evaluating both the performance and scalability of the proposed model in a realistic quantum computational setting. In particular, the use of a quantum annealer such as the D-Wave system will allow us to explore the advantages and limitations of quantum optimization when applied to high-dimensional energy landscapes associated with molecular docking. A central objective of this next phase will be to assess whether the quantum annealer is capable of consistently identifying low-energy configurations comparable to those obtained via classical simulated annealing, especially in the presence of a larger number of binary variables and more intricate molecular geometries. We plan to apply the model to a broader class of ligands and binding pockets, including larger aromatic systems and flexible molecules, in order to test the generalizability of the Hamiltonian formulation and the robustness of the optimization pipeline. These future developments will help clarify the practical role of quantum annealing in computational chemistry workflows and contribute to the development of hybrid classical-quantum methods for molecular modeling. In the coming months, we also plan to carry out the calculations related to the T-shaped configuration, which have not yet been reported. Furthermore, to further validate the proposed model, we are interested in conducting new simulations aimed at investigating the energy profile associated with the spatial rotations of the benzenic fragment. Another key challenge that emerged concerns the trade-off between model accuracy and computational cost. Unlike the formulation in [1], the model presented in this thesis explicitly includes hydrogen atoms in the optimization process.

While this choice improves the level of chemical detail, it also significantly increases the number of variables involved, making the annealing procedure more computationally demanding. A potential future direction will therefore be the search for simplified yet chemically consistent formulations that retain predictive power while reducing resource consumption.

In conclusion, this work shows that discrete Hamiltonians in QUBO form, when grounded in physical data and carefully structured, offer a promising mathematical framework for approximating molecular interactions. From a mathematical standpoint, this thesis demonstrates the potential of combining graph-based modeling, empirical parametrization, and discrete optimization techniques to address relevant challenges in molecular docking and computational chemistry more broadly.

Bibliography

[1] Emanuele Triuzzi, Riccardo Mengoni, Francesco Micucci, Domenico Bonanni, Daniele Ottaviani, Andrea Beccari, and Gianluca Palermo. Molecular docking via weighted subgraph isomorphism on quantum annealers. *arXiv preprint arXiv:2405.06657*, 2024.

[2] Abraham P Punnen. The quadratic unconstrained binary optimization problem. *Springer International Publishing*, 10:978–3, 2022.

[3] Gary Kochenberger, Jin-Kao Hao, Fred Glover, Mark Lewis, Zhipeng Lü, Haibo Wang, and Yang Wang. The unconstrained binary quadratic programming problem: a survey. *Journal of combinatorial optimization*, 28:58–81, 2014.

[4] Scott Kirkpatrick, C Daniel Gelatt Jr, and Mario P Vecchi. Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[5] Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[6] Martin Pincus. A monte carlo method for the approximate solution of certain types of constrained optimization problems. *Operations research*, 18(6):1225–1228, 1970.

[7] A Khachaturyan, S Semenovskaya, and B Vainstein. Statistical-thermodynamic approach to determination of structure amplitude phases. *Sov. Phys. Crystallography*, 24(5):519–524, 1979.

[8] A Khachaturyan, S Semenovsovskaya, and B Vainshtein. The thermodynamic approach to the structure analysis of crystals. *Foundations of Crystallography*, 37(5):742–754, 1981.

[9] Davide Pastorello et al. *Concise guide to quantum machine learning.* Springer, 2023.

[10] Arnab Das and Bikas K Chakrabarti. *Quantum annealing and related optimization methods*, volume 679. Springer Science & Business Media, 2005.

[11] Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse ising model. *Physical Review E*, 58(5):5355, 1998.

[12] Satoshi Morita and Hidetoshi Nishimori. Mathematical foundation of quantum annealing. *Journal of Mathematical Physics*, 49(12), 2008.

[13] D-Wave Systems Inc. Simulated annealing sampler. `https://dwave-neal-docs.readthedocs.io/en/latest/reference/sampler.html`, 2025. Accessed: 2025-07-11.

[14] D-Wave Systems Inc. Topologies. `https://docs.dwavequantum.com/en/latest/quantum_research/topologies.html#topology-intro-pegasus`, 2025. Accessed: 2025-07-11.

[15] A Sharma and Pradip Maharjan. Quantum annealing as an optimized simulated annealing: a case study, 2018.

[16] Giuseppe E Santoro and Erio Tosatti. Optimization using quantum mechanics: quantum annealing through adiabatic evolution. *Journal of Physics A: Mathematical and General*, 39(36):R393, 2006.

[17] Leonardo Banchi, Mark Fingerhuth, Tomas Babej, Christopher Ing, and Juan Miguel Arrazola. Molecular docking with gaussian boson sampling. *Science advances*, 6(23):eaax1950, 2020.

[18] Stefano Mensa, Emre Sahin, Francesco Tacchino, Panagiotis Kl Barkoutsos, and Ivano Tavernelli. Quantum machine learning framework for virtual screening in drug discovery: a prospective quantum advantage. *Machine Learning: Science and Technology*, 4(1):015023, 2023.

[19] Kevin Mato, Riccardo Mengoni, Daniele Ottaviani, and Gianluca Palermo. Quantum molecular unfolding. *Quantum science and technology*, 7(3):035020, 2022.

[20] G Patrick Brady and Pieter FW Stouten. Fast prediction and visualization of protein binding pockets with pass. *Journal of computer-aided molecular design*, 14:383–401, 2000.

[21] Jinyin Zha, Jiaqi Su, Tiange Li, Chongyu Cao, Yin Ma, Hai Wei, Zhiguo Huang, Ling Qian, Kai Wen, and Jian Zhang. Encoding molecular docking for quantum computers. *Journal of Chemical Theory and Computation*, 19(24):9018–9024, 2023.

[22] T Langer, RD Hoffmann, F Bachmair, and S Begle. Chemical function based pharmacophore models as suitable filters for virtual 3d-database screening. *Journal of Molecular Structure: THEOCHEM*, 503(1-2):59–72, 2000.

[23] Brandon Schramm, Montgomery Gray, and John M Herbert. Substituent and heteroatom effects on π–π interactions: Evidence that parallel-displaced π-stacking is not driven by quadrupolar electrostatics. *Journal of the American Chemical Society*, 2025.

[24] Christopher A Hunter and Jeremy KM Sanders. The nature of. pi.-. pi. interactions. *Journal of the American Chemical Society*, 112(14):5525–5534, 1990.

[25] Hans-Jörg Schneider, Klaus Philippi, and Jürgen Pöhlmann. Structure- and solvent-dependence in the complexation of lipophilic substrates in a water-soluble azacyclophane. *Angewandte Chemie International Edition in English*, 23(11):908–910, 1984.

[26] Ben Askew, Pablo Ballester, Chris Buhr, Kyu Sung Jeong, Sharon Jones, Kevin Parris, Kevin Williams, and Julius Rebek Jr. Molecular recognition with convergent functional groups. vi. synthetic and structural studies with a model receptor for nucleic acid components. *Journal of the American Chemical Society*, 111(3):1082–1090, 1989.

[27] Alexander V Muehldorf, Donna Van Engen, John C Warner, and Andrew D Hamilton. Aromatic-aromatic interactions in molecular recognition: a family of artificial receptors for thymine that shows both face-to-face and edge-to-face orientations. *Journal of the American Chemical Society*, 110(19):6561–6562, 1988.

[28] Marcel Swart, Tushar van der Wijst, Célia Fonseca Guerra, and F Matthias Bickelhaupt. π-π stacking tackled with density functional theory. *Journal of molecular modeling*, 13:1245–1257, 2007.

[29] Kevin Carter-Fenk and John M Herbert. Electrostatics does not dictate the slip-stacked arrangement of aromatic π–π interactions. *Chemical Science*, 11(26):6758–6765, 2020.

[30] KI Ramachandran, Deepa Gopakumar, and Krishnan Namboori. *Computational chemistry and molecular modeling: principles and applications*. Springer, 2008.

[31] Paul Geerlings, Frank De Proft, and Wilfried Langenaeker. Conceptual density functional theory. *Chemical reviews*, 103(5):1793–1874, 2003.

[32] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[33] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[34] Douglas R Hartree. The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 89–110. Cambridge university press, 1928.

[35] Vladimir Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61(1):126–148, 1930.

[36] Krzysztof Szalewicz. Symmetry-adapted perturbation theory of intermolecular forces. *Wiley interdisciplinary reviews: computational molecular science*, 2(2):254–272, 2012.

[37] John C Slater. The theory of complex spectra. *Physical review*, 34(10):1293, 1929.

[38] Paolo Giannozzi, Oliviero Andreussi, Thomas Brumme, Oana Bunau, M Buongiorno Nardelli, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Matteo Cococcioni, et al. Advanced capabilities for materials modelling with quantum espresso. *Journal of physics: Condensed matter*, 29(46):465901, 2017.

[39] Q-Chem, Inc. Chapter 12. section s10. subsection s4: Sapt data collection. `https://manual.q-chem.com/latest/Ch12.S10.SS4.html`, 2025. Accessed: 2025-07-09.

[40] Thomas A Halgren. Merck molecular force field. ii. mmff94 van der waals and electrostatic parameters for intermolecular interactions. *Journal of Computational Chemistry*, 17(5-6):520–552, 1996.

[41] Jinfeng Shao, Bastiaan P Kuiper, Andy-Mark WH Thunnissen, Robbert H Cool, Liang Zhou, Chenxi Huang, Bauke W Dijkstra, and Jaap Broos. The role of tryptophan in π interactions in proteins: An

experimental approach. *Journal of the American Chemical Society*, 144(30):13815–13822, 2022.

[42] SK Burley and Gregory A Petsko. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. *Science*, 229(4708):23–28, 1985.