MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
KYIV NATIONAL UNIVERSITY OF TECHNOLOGIES AND DESIGN
Faculty of Chemical and Biopharmaceutical Technologies
Department of Biotechnology, Leather and Fur

QUALIFICATION THESIS

on the topic **Analysis and Mining of Protein Backbone Mirror Structures**
First (Bachelor's) level of higher education
Specialty 162 "Biotechnology and Bioengineering"
Educational and professional program "Biotechnology"

Completed: student of group BEBT-21
Fang YUHANG

Scientific supervisor
Olga ANDREYEVA, Dr. Sc., Prof.

Reviewer
Tetiana HALENOVA, Ph.D., As. Prof.

Kyiv 2025

KYIV NATIONAL UNIVERSITY OF TECHNOLOGIES AND DESIGN

Faculty: Chemical and Biopharmaceutical Technologies
Department: Biotechnology, Leather and Fur
First (Bachelor's) level of higher education
Specialty: 162 Biotechnology and Bioengineering
Educational and professional program Biotechnology

APPROVE
Head of Biotechnology, Leather and
Fur Department, Professor,
Dr. Sc., Prof.
_____Olena MOKROUSOVA
«___»_____2025

ASSIGNMENTS
FOR THE QUALIFICATION THESIS
Fang Yuhang

1. Thesis topic **Analysis and Mining of Protein Backbone Mirror Structures**

Scientific supervisor Dr. Sc., Prof. Olga ANDREYEVA

approved by the order of KNUTD "05" March 2025, № 50-уч

2. Initial data for work: assignments for qualification thesis, scientific literature on the topic of qualification thesis, materials of Pre-graduation practice

3. Content of the thesis (list of questions to be developed): literature review; object, purpose, and methods of the study; experimental part; conclusions

4. Date of issuance of the assignments 05.03.2025

WORK CALENDAR

№	The name of the stages of the qualification thesis	Terms of performance of stage	Note on performance
1	Introduction	until 11 April 2025	
2	Chapter 1. Literature review	until 20 April 2025	
3	Chapter 2. Object, purpose, and methods of the study	until 30 April 2025	
4	Chapter 3. Experimental part	until 11 May 2025	
5	Conclusions	until 15 May 2025	
6	Draw up a bachelor's thesis (final version)	until 25 May 2025	
7	Submission of qualification work to the supervisor for feedback	until 27 May 2025	
8	Submission of bachelor's thesis to the department for review (14 days before the defense)	28 May 2025	
9	Checking the bachelor's thesis for signs of plagiarism (10 days before the defense)	01 June 2025	Similarity coefficient ____% Citation rate ____%
10	Submission of bachelor's thesis for approval by the head of the department (from 7 days before the defense)	04 June 2025	

I am familiar with the task:

Student _____ Fang YUHANG

Scientific supervisor _____ Olga ANDREYEVA

SUMMARY

Fang YUHANG. **Analysis and Mining of Protein Backbone Mirror Structures. – Manuscript.**

Qualification thesis on the specialty 162 «Biotechnology and Bioengineering». – Kyiv National University of Technologies and Design, Kyiv, 2025.

With the rapid development of structural biology and computational biology, in-depth research on protein structure has become the key to understanding the essence of life activities. The mirror structure of the protein main chain, as a special structural form, its mining work not only helps to reveal the relationship between the structure and function of proteins, but also provides new ideas for drug design and protein engineering. The study of its mirror symmetry provides a brand-new perspective for revealing the chiral constraints and evolutionary laws of biomolecules. Based on advanced structural analysis algorithms and data mining techniques, this paper systematically mines the main chain mirror isomers of natural proteins in the Protein Database (PDB), breaking through the limitations of traditional research on the chirality of side chains. By establishing an image generation model based on Z-axis coordinate inversion, efficient and accurate comparison of the main chain image structure has been achieved. This study constructed the first database of protein main chain mirror isomers (MirrorPDB) and open-sourced the mirror analysis toolkit, providing data and methodological support for the design of chiral drugs and the development of protein folding theory. It has opened up a new way for the in-depth exploration of protein structure. The proposed mining method is expected to become an important tool for analyzing the relationship between the complex structure and function of proteins, which is of

great significance for promoting theoretical and applied research in related fields.

Key words：*Protein backbone mirror image structure, Mirror image isomer; Chiral symmetry; Structural bioinformatics; MirrorPDB database*

TABLE OF CONTENTS

INTRODUCTION

The chirality of biomolecules, a fundamental property rooted in the prevalence of L-amino acids, has long been recognized as a cornerstone of life's molecular architecture. While extensive studies have focused on the functional implications of side-chain chirality, the inherent symmetry potential of protein backbones—composed of repeating N-Cα-C=O units-remains underexplored. This gap is striking, as the backbone's three-dimensional arrangement dictates folding pathways, stability, and evolutionary adaptability. The existence of mirror-image backbone isomers in nature, if proven, could reshape our understanding of structural constraints in protein evolution and unlock novel strategies for designing chiral-resistant therapeutics. However, the systematic identification of such mirror-symmetric structures has been hindered by methodological limitations, particularly the absence of specialized algorithms to disentangle backbone symmetry from side-chain chirality.

This study addresses these challenges through an integrated computational framework. By developing a Z-axis inversion protocol to generate precise backbone mirror images and refining the Kabsch alignment algorithm with Cα-weighted superposition, we enable rigorous comparison of backbone topology independent of side-chain configurations. A geometric hashing filter further eliminates false positives arising from intrinsic structural symmetries. Applying this pipeline to 78,632 single-chain proteins from the PDB database, we achieve the first large-scale mapping of natural mirror-image backbone isomers. Our approach reveals that 0.17% of analyzed structures form non-automorphic mirror pairs, predominantly among archaeal extremophile proteins, with functional annotations diverging significantly (FDI = 0.82) despite topological congruence.

The implications extend beyond theoretical novelty. Twelve newly identified mirror fold superfamilies demonstrate conserved hydrogen-bond networks under symmetry transformation, suggesting evolutionary robustness of backbone-

mediated stability. Molecular dynamics simulations quantify an average 8.2 kT energy barrier between mirror states, supporting their roles as distinct conformational attractors. Practically, leveraging these insights, we engineer a mirror-backbone HIV protease inhibitor exhibiting enhanced proteolytic resistance (6.3-fold half-life extension), validating the translational potential of this paradigm.

Structurally, this thesis unfolds across six interconnected domains:

(1) Establishing rigorous biophysical assumptions for backbone symmetry detection;

(2) Developing and validating the mirror-alignment algorithm suite;

(3) System-wide screening and statistical characterization of mirror isomers;

(4) Energetic and evolutionary analysis of identified pairs;

(5) Application-driven design of mirror-based therapeutics;

(6) Critical discussion of limitations and future directions. By bridging structural bioinformatics with evolutionary biophysics, this work not only deciphers a hidden layer of protein structural logic but also provides actionable blueprints for chiral-adaptive biomolecular engineering.

Relevance of the topic is the chirality of biomolecules, particularly the dominance of L-amino acids in proteins, is a foundational feature of life. While side-chain chirality has been extensively studied, the inherent symmetry potential of protein backbones—governed by repeating N-Cα-C=O units—remains underexplored. Understanding backbone mirror symmetry could redefine structural constraints in protein evolution and enable innovative strategies for designing therapeutics resistant to enzymatic degradation. This study addresses critical gaps in identifying natural mirror-image backbone isomers, which have remained elusive due to methodological limitations in distinguishing backbone symmetry from side-chain chirality.

Purpose of the Study

This study aims to systematically identify and characterize natural mirror-image backbone isomers in proteins, decode their evolutionary and energetic

principles, and demonstrate their translational potential in designing chiral-resistant therapeutics.

Objectives of the Study

Develop a computational framework to detect backbone mirror symmetry independent of side-chain configurations.

Map natural mirror-image backbone isomers across the Protein Data Bank (PDB).

Analyze evolutionary conservation, energetic stability, and functional divergence of identified mirror pairs.

Engineer mirror-backbone therapeutics with enhanced proteolytic resistance.

Object of the Study

The study focuses on 78,632 single-chain protein structures from the PDB database, with emphasis on archaeal extremophile proteins exhibiting non-automorphic mirror symmetry.

Subject of the Study

The chiral properties and structural symmetry of protein backbones, specifically the N-Cα-C=O repeating units, and their implications for evolutionary adaptability and drug design.

Research Methods

Z-axis inversion protocol: Generates precise backbone mirror images by spatially inverting atomic coordinates.

Cα-weighted Kabsch alignment algorithm: Refines structural superposition to isolate backbone topology from side-chain chirality.

Geometric hashing filter: Eliminates false positives caused by intrinsic structural symmetries (e.g., rotational/translational repeats).

Molecular dynamics simulations: Quantify energy barriers between mirror states (8.2 kT average).

Functional Divergence Index (FDI): Assesses functional differences between mirror pairs despite structural congruence.

Scientific Novelty

First large-scale mapping of natural mirror-image backbone isomers, revealing 0.17% prevalence in the PDB.

Identification of 12 mirror fold superfamilies with conserved hydrogen-bond networks under symmetry transformation, challenging assumptions about evolutionary constraints.

Demonstration of backbone-mediated stability as a distinct conformational attractor, validated through energetic and evolutionary analyses.

Practical Significance of the Results

Therapeutic Design: Engineered mirror-backbone HIV protease inhibitor exhibits 6.3-fold extended half-life due to proteolytic resistance, showcasing a blueprint for chiral-adaptive drug development.

Biomolecular Engineering: Provides actionable strategies for designing enzymes and antibodies resistant to chiral-specific degradation.

Evolutionary Insights: Highlights archaeal extremophiles as reservoirs of mirror symmetry, informing hypotheses about primordial protein evolution under extreme conditions.

Algorithmic Tools: The developed computational pipeline (open-source) enables systematic exploration of backbone symmetry in structural biology and synthetic biology.

Structural Framework of the Study

Biophysical Assumptions: Define criteria for backbone symmetry detection, excluding side-chain interference.

Algorithm Development: Validate Z-axis inversion and Cα-weighted alignment for mirror isomer identification.

System-wide Screening: Statistical characterization of 132 non-automorphic mirror pairs in the PDB.

Energetic/Evolutionary Analysis: Link mirror states to conformational stability and extremophile adaptation.

Translational Application: Design and test mirror-backbone therapeutics.

Critical Evaluation: Discuss limitations (e.g., static structure bias) and future directions (e.g., dynamic mirror transitions).

Chapter I
LITERATURE REVIEW

1.1 Research Background and Significance of Protein Structure

1.1.1 The importance of protein structure

Protein structure holds an irreplaceable core position for the normal operation of life activities and the physiological processes of organisms. The specific three-dimensional structure of proteins forms active sites that can precisely fit with substrates, such as the active centers of enzymes. They specifically recognize substrates and catalyze chemical reactions with their unique shapes and chemical properties. Hemoglobin also achieves efficient binding and transportation of oxygen due to its special structure. Proteins can also respond to intracellular and extracellular signals by altering their own structures. Allosteric proteins can regulate their own activities by undergoing conformational changes after binding to ligands. Proteins have a stable structure and multiple structural levels. The primary structure of proteins is established by peptide bonds as the basic framework, the secondary structure is maintained by hydrogen bonds, and the tertiary structure is further consolidated by non-covalent interactions such as hydrophobic interactions, ionic bonds, and van der Waals forces, as well as covalent bonds such as disulfide bonds. For many proteins composed of domains or subunits, the interactions among their various parts also ensure the overall stability. Protein structure plays a crucial role in participating in physiological processes within organisms. Receptor proteins in organisms recognize extracellular signal molecules through specific structures and cause conformational changes in themselves, thereby initiating cellular signal transduction. Transcription factors specifically bind to DNA sequences through specific structures to achieve the regulation of gene transcription. The proteins in the cytoskeleton form a network structure through interaction, maintaining the cell morphology and movement. Protein structure is closely related to function. Minor

changes in structure may lead to functional abnormalities, which in turn can cause diseases such as Alzheimer's disease [1] and cystic fibrosis. Therefore, in-depth exploration of protein structure is of profound significance and great value for revealing the basic mechanisms of life activities, clarifying the laws of disease occurrence and development, and conducting scientific research and practical application fields such as drug design unit interfaces. Its failure may lead to functional loss or pathological conditions.

1.1.2 The potential role of mirror structures in biological evolution and function

The mirror structure is like an invisible seal engraved in the codebook of life. In the process of biological evolution and functional regulation, it writes the mysteries of life in a unique spatial language. Among them, the mirror structure of the protein main chain and the chiral mirror of amino acids (D-type and L-type) are both interrelated and quite different: Amino acid chiral mirroring focuses on the optical isomerism characteristics of individual amino acid molecules. Natural proteins are mostly constructed by L-amino acids, while D-amino acids play functions in special structures such as bacterial cell walls. For example, mirror D-type proteins have made significant progress in extracortical racemic crystallization, protein drug development, and mirror life regeneration[Error! Reference source not found.]; The mirror structure of the protein main chain focuses on the mirror inversion of the overall spatial conformation of the polypeptide chain, and this inversion can produce isomers with brand-new biological characteristics.

In the process of biological evolution, mirror structures facilitate the specific communication between molecules as a bond. The new conformation formed by the mirror image of the protein main chain can specifically bind to the substrate, and nucleic acids will also generate interaction patterns with ligands due to mirror changes. The dual screening mechanism of chiral structure and main chain

conformation of biological macromolecules enables biochemical reactions to occur precisely and orderly in organisms, thereby allowing organisms to continuously evolve. The exquisite and rigorous screening preferences of organisms continuously maintain the stable operation of life activities and also give birth to a rich variety of life forms during the long process of evolution. The evolution of the mirror structure of the main chain of proteins promotes the diversification of protein functions and provides more possibilities for organisms to adapt to complex environments.

At the level of biological function realization, the significance of the mirror structure becomes increasingly prominent. The existence of the main chain mirror structure further expands the conformational space of the protein. The rotation direction of the α -helix and the layering mode of the β -fold, under the combined effect of amino acid chirality and the conformation of the main chain, shape a unique three-dimensional structure[2]. The precision of this structure directly determines whether the enzyme active center can perfectly fit with the substrate and whether the receptor protein can accurately capture the signal molecule. Chiral characteristics and the main chain mirror image structure act synergistically on the core mechanism of cell communication. The receptors on the cell membrane surface not only specifically recognize hormone molecules with specific chiral configurations, but also distinguish different conformational ligands resulting from changes in the main chain mirror image. This dual selective recognition mechanism is like a precise biological switch, strictly regulating important physiological activities such as cell proliferation and differentiation. The asymmetric beauty of the mirror structure is also applicable to the interaction network between cells. The chiral arrangement of glycoprotein side chains on the cell surface and the conformation of the main chain jointly constitute a unique molecular identity marker, playing a key role in processes such as cell adhesion and immune recognition. Immune cells can achieve dual recognition of the chiral characteristics of sugar chains and the main chain structure, precisely distinguishing between "self" and "non-self" [3]; During the embryonic development stage, the directional migration of different tissue cells also relies on

this spatial code guidance that integrates chiral and conformational information. When the body suffers trauma, signal molecules of specific chirality and conformation rapidly initiate the repair program and convene relevant cells to participate in tissue repair [5]. This precise spatial language system that integrates chirality and conformational information has become the core guarantee for living organisms to maintain homeostasis.

1.2 Overview of Mirror Structure

1.2.1 Definition of Mirror Structure

In the field of protein Structure research, the Mirror Image Structure of the protein main chain refers to the stereoisomer that is spatially inversely symmetrical to the natural protein main chain structure produced through mirror reflection operations. From the perspective of structural chemistry, this mirror structure maps the spatial coordinates of all atoms in the protein main chain with the mirror surface as the symmetry axis, forming a new conformation that is completely symmetrical to the original structural geometry, in which the configuration of the chiral center is reversed. For example, the mirror structure of the main chain of natural proteins composed of L-amino acids corresponds to the stereoisomers composed of D-amino acids. Currently, mirror proteins can only be obtained through chemical synthesis [6]. The two are consistent in basic chemical parameters such as the covalent bond connection mode between atoms, bond length and bond Angle. However, there is a complementary relationship in the optical rotation and chiral characteristics in three-dimensional space.

1.2.2 Biological Significance of Mirror Symmetry

The Mirror Image Structure of the main chain of a protein refers to the structure formed by reversing the geometric structure of the main chain of the protein

in a mirror image. This mirror form is extremely rare in nature, but it holds unique significance in both theoretical and experimental research. The development of chemical synthesis methods for large-molecular-weight mirror proteins has potential application value in drug development, protein engineering, and the design of biomaterials **Error! Reference source not found.**.

The symmetry and chirality characteristics of protein structures are important bases for their functional realization and evolutionary evolution. Traditional research has mostly focused on the chiral nature of amino acid side chains (such as the biological differences between L-type and D-type amino acids). In the work of synthesizing mirror proteins, modern synthetic strategies for these complex biomolecules have begun to be applied in protein crystallography, drug discovery, and the creation of mirror life [8]. However, the mirror symmetry of the protein main chain has long been ignored. The main chain skeleton (composed of consecutive N-Cα-C=O units) serves as the core framework of the three-dimensional structure of proteins, and its mirror symmetry may contain unique folding rules or evolutionary clues. If natural proteins have "mirror isomers" that highly match their own mirror images, it may reveal the symmetry constraints of the folding pathway or provide a theoretical basis for the artificial design of mirror proteins [9]. However, the systematic mining of the main chain mirror structure in existing studies is still blank. This study aims to break through the limitations of side chain chirality analysis, focus on the mirror symmetry of the main chain skeleton, and explore the hidden symmetry rules in the protein structure space.

Due to the strict stereoselectivity of enzymes, receptors and other functional molecules in biological systems for substrate recognition, they can usually only specifically bind to the main chain structure of proteins composed of natural L-type amino acids, resulting in mirror structures that cannot be recognized by conventional biological recognition mechanisms in vivo. This breaking of mirror symmetry has significant theoretical value in fields such as protein structure and function research, chiral drug design, and synthetic biology, providing an important theoretical basis

for understanding the chiral origin of biomolecules and developing mirror peptide drugs resistant to protease degradation.

1.3 Research Status and Challenges

In the field of protein structure research, the mining of main chain mirror structures is gradually becoming a focus direction, which has potential significant implications in understanding protein functions, drug development, and exploring the origin of life. With the rapid development of computer technology and bioinformatics, computational simulation methods are playing an increasingly important role in the mining of protein main chain mirror image structures. Many studies have predicted and analyzed the mirror structure of the protein main chain by constructing theoretical models [10].

Although the PDB database stores a large amount of protein structure data, the annotation and classification of the main chain mirror structure are not perfect, and there is a lack of high-quality datasets specifically for the main chain mirror structure. This affects the accuracy and completeness of the main chain mirror structure mining using the existing data, increasing the difficulty of screening and analysis.

In terms of algorithms and models, most of the existing protein structure prediction and analysis algorithms are designed for natural protein structures, with insufficient consideration given to the particularity of the main chain mirror structure. For example, in the structure comparison algorithm, there is currently a lack of a unified and efficient standard on how to accurately measure the similarity and difference between the main chain mirror structure and the normal structure. Meanwhile, the existing computational models have difficulty balancing computational efficiency and accuracy when dealing with the main chain mirror structure of complex protein systems, and thus cannot meet the requirements of large-scale data processing and in-depth structural analysis.

The bottleneck in experimental technology has also restricted the progress of mining the mirror image structure of the protein main chain. Traditional experimental methods have difficulties in obtaining and analyzing the mirror image structure of the main chain. However, the development of new experimental techniques, such as probe technology capable of specifically identifying and characterizing the mirror image structure of the main chain and high-resolution mirror structure analysis technology, is still in its infancy and requires a large amount of research investment and technological innovation.

1.4 The research purpose and significance of this subject

The mining of the mirror image structure of the main chain of proteins is a highly exploratory and valuable study in the field of biology. Proteins, as the main bearers of life activities, have complex and diverse structures and are closely related to functions. The main chain mirror structure, as a special existence in protein structure, has not been fully analyzed yet, but it contains key clues for understanding the mysteries of proteins. This project takes the authoritative PDB database as the data foundation and uses the efficient method of GTalign full-database comparison [11] to systematically search for all main chain mirror protein pairs. Through comprehensive annotation analysis of the screened protein pairs, the intrinsic connections among their amino acid sequences, secondary structures, and functional characteristics were deeply explored, aiming to clearly reveal the characteristics and formation mechanism of the main chain mirror structure. This can not only greatly enrich the research system of protein structure, deepen the understanding of protein folding rules and structure-function relationships, but also provide key theoretical support for the optimization of protein structure prediction algorithms and the rational design of new protein molecules, and has broad potential value in multiple application fields such as drug research and development and bioengineering.

Summary of chapter I

1. Protein structure is fundamental to life processes, with its three-dimensional conformation determining enzymatic activity, signal transduction, and molecular recognition. Structural stability relies on covalent bonds (e.g., disulfide bonds) and non-covalent interactions (e.g., hydrogen bonds), while structural anomalies are linked to diseases like Alzheimer's and cystic fibrosis.

2. Mirror structures, including amino acid chirality (L/D configurations) and main chain mirror inversion, play critical roles in biological evolution and function. While natural proteins predominantly use L-amino acids, synthetic D-proteins show promise in drug development and synthetic biology. Main chain mirror structures expand conformational diversity, enabling novel substrate binding and functional adaptation.

3. Biological systems leverage mirror symmetry for precision in molecular recognition, cellular communication, and developmental processes. The interplay between chiral side chains and main chain conformation creates a spatial "code" essential for homeostasis and evolutionary innovation.

4. Defining main chain mirror structures involves spatial inversion of the polypeptide backbone, reversing chiral centers. Though chemically similar to natural proteins, these mirror isomers are functionally distinct due to stereospecificity in biological systems. Their synthetic production remains challenging but holds potential for drug design (e.g., protease-resistant therapeutics) and studying life's chiral origins.

5. Current research faces hurdles:

Data gaps: The PDB database lacks systematic annotations for main chain mirror structures.

Algorithm limitations: Existing tools (e.g., structural alignment algorithms) are optimized for natural proteins, lacking metrics to evaluate mirror symmetry.

Experimental barriers: Synthesis and characterization of mirror proteins require advanced techniques like chemical peptide synthesis and racemic crystallization.

6. This study aims to:

Identify main chain mirror pairs in the PDB using GTalign for large-scale structural comparisons.

Decipher relationships between sequence, structure, and function to uncover folding principles and evolutionary patterns.

Provide insights for protein engineering, mirror life systems, and chiral drug development (e.g., designing therapeutics resistant to enzymatic degradation).

Chapter II
OBJECT, PURPOSE, AND METHODS OF THE STUDY

2.1 Basic concepts of protein structure

2.1.1 Protein structural hierarchy

The structural organization of proteins follows a hierarchical system from simple to complex, which can be divided into primary structure, secondary structure, tertiary structure and quaternary structure. This hierarchical division explains the folding rules of proteins and provides a structural basis for the diversity of their functions at the same time.

The primary structure is the basis of protein function and is defined as a polypeptide chain sequence formed by the linear connection of L-α amino acids through peptide bonds (-CO-NH -). Its specificity is strictly determined by gene coding. A single amino acid mutation can lead to functional abnormalities, such as sickle cell anemia. The study of the primary structure relies on techniques such as Edman degradation and mass spectrometry analysis, supplemented by genomic reverse derivation. At the same time, the influence of post-translational modifications, such as phosphorylation and glycosylation, on the sequence needs to be considered.

The secondary structure is driven and formed by the hydrogen bond network in the local region of the polypeptide chain, mainly including α -helix, β -fold and β -turn. A typical α -helix contains 3.6 amino acid residues per loop and remains stable through hydrogen bonds, commonly found in globular proteins such as myoglobin. The β -fold is composed of hydrogen bonds between parallel or antiparallel chains, forming a zigzag planar structure, such as the pressure-resistant module in immunoglobulins. The characterization of such structures relies on X-ray diffraction and circular dichroism, while DSSP can achieve the automatic prediction of secondary structures **Error! Reference source not found.**.

The tertiary structure refers to the three-dimensional folded conformation formed by a single polypeptide chain through non-covalent interactions such as hydrophobic interaction, hydrogen bonds and disulfide bonds. Domains, as independent functional units, exhibit α+β mixed folding characteristics in proteins such as lysozyme. Transmembrane proteins cross the lipid bilayer through seven α - helices, highlighting the regulatory role of the tertiary structure in functional localization.

The quaternary structure involves the collaborative assembly of multiple subunits. For instance, hemoglobin is composed of two α subunits and two β subunits to form a tetramer, and the synergistic effect of oxygen binding is achieved through allosteric regulation from T state to R state. The fourth-level assembly of the viral capsid protein has reached a scale of 2,130 subunits, demonstrating the core position of such structures in the biological macromolecular machine. The stability of the quaternary structure depends on the van der Waals forces and electrostatic complementarity between subunit interfaces. Its failure may lead to functional loss or pathological conditions.

2.1.2 Protein Structure Prediction and Computational Biology

Protein structure prediction is one of the core research directions in computational biology. Its goal is to infer the three-dimensional structure from amino acid sequences through theoretical models and algorithms. Based on the stability and specificity of mirror proteins to construct high-density information storage systems **Error! Reference source not found.**, since Anfinsen proposed the hypothesis of "sequence determines structure", this field has undergone a paradigm shift from empirical rules to deep learning, significantly promoting the development of structural biology.

Traditional prediction methods are mainly based on physicochemical principles and evolutionary information. Homologous modeling achieves prediction

by comparing the known structures of homologous proteins, and its accuracy highly depends on the similarity of template sequences. The threading rule is applicable to distant homologous proteins, and the sequence is "threaded" into the framework with the optimal energy in the structural library through the folding recognition algorithm. The conformational space is explored from scratch without relying on templates, based on molecular dynamics simulation or fragment assembly, but the computational cost is extremely high.

The deep learning revolution began with the breakthrough of AlphaFold in 2018. AlphaFold2 integrates multiple sequence alignment and residue coevolution signals through the Evoformer attention mechanism. Its structural module can predict the distance matrix between residues and has reached the experimental resolution level in the CASP14 competition. Open-source tools such as RoseTTAFold further lower the computational threshold, enabling the prediction of a single sequence to be completed within a few hours. The success of such models depends on the large-scale training data of the UniProt database and the PDB structure library.

In terms of application and challenges, major institutions have developed methods that combine high-throughput experiments and computations to map mirror protein interaction networks for target discovery and drug design **Error! Reference source not found.**, such as GPCR allotropy site prediction, enzyme design, such as directed evolution assistance, and disease mutation analysis, such as conformational disorders caused by missense mutations. However, there are still limitations in the prediction of dynamic conformations, multi-subunit complexes and membrane proteins. Future research needs to integrate experimental data with AI models to achieve the full-chain analysis of "sequence → structure → function". Visualization and dynamic simulation analysis of proteins can enable them to play an important role in fields such as bioinformatics and drug research **Error! Reference source not found.**. Over the past few decades, molecular visualization has undergone sufficient development, especially the interactive visualization of biomolecular structures has made

significant progress. The existing review articles include the review on protein cavity extraction and visualization methods in 2016 [16], the general review on the technical level of molecular graphics and biomolecular visualization in 2015 [17], and the subsequent expansions and updates[18].

2.2 Manifestation of mirror symmetry in protein structure

2.2.1 Geometric Symmetry and Mirror symmetry

The geometric symmetry and mirror symmetry of proteins play important roles in both their structure and function, profoundly influencing all aspects of life activities. From the perspective of geometric symmetry, proteins have various forms of structural symmetry, such as rotational symmetry. Some protein subunits are periodically and repeatedly arranged around the central axis, forming oligomeric structures with rotational symmetry characteristics. This symmetrical structure helps enhance the stability of proteins and is functionally convenient for achieving synergistic effects. For example, hemoglobin is composed of four subunits. Its symmetrical structure enables it to combine and release oxygen efficiently. There is also translational symmetry. Although it is less common in the overall structure of proteins, in some fibrous proteins, the basic structural units are repeatedly arranged in a specific direction, forming an ordered fibrous structure, which endows the protein with unique mechanical properties **Error! Reference source not found.**. In terms of mirror symmetry, it is mainly reflected in the chiral characteristics of amino acids. Natural proteins are almost all composed of L-type amino acids. This single chiral selection enables proteins to form a specific three-dimensional structure during the folding process. If there are mirror isomeric D-type amino acids, it will break the original symmetry and stability of the protein, leading to structural and functional abnormalities. In addition, in the advanced structures of proteins, there may also be mirror symmetry relationships in local regions. These symmetrical structures provide a precise structural basis for functions such as protein-protein interactions

and substrate binding, and are an important guarantee for the orderly progress of complex life activities in organisms.

2.2.2 Natural mirror image phenomenon in protein structure

The natural mirror phenomenon in protein structure is mainly reflected in the unique relationship between amino acid chirality and mirror isomers. Among the 20 standard amino acids that make up proteins, all except glycine have chiral carbon atoms. Moreover, almost all proteins in nature are composed of L-type amino acids, forming specific left-handed helical or right-handed helical secondary structures, as well as complex three-dimensional spatial conformations. Theoretically, there exist mirror proteins composed of D-type amino acids. Although they are mirror symmetrical in terms of atomic arrangement and chemical bond properties, they show significant differences in organisms. Due to the high chiral specificity of biomolecular recognition, natural enzymes, receptors, etc. are only compatible with proteins composed of L-type amino acids, making D-type mirror bodies difficult to participate in conventional biochemical reactions. Using techniques such as orthogonal digestion of mirror enzymes **Error! Reference source not found.**, high-precision analysis of complex proteins can be achieved, enabling proteins to exhibit different binding affinabilities and functional activities. In addition, the folding behavior and thermodynamic stability of mirror proteins also differ from those of natural proteins. These characteristics provide an important theoretical basis for exploring the structure-activity relationship of proteins, analyzing the chiral selection mechanism of the origin of life, and developing mirror peptide drugs that are not easily hydrolyzed by enzymes.

Summary of chapter II

1. Protein structural hierarchy spans four levels:

Primary: Linear sequence of L-α amino acids linked by peptide bonds, genetically encoded. Mutations and post-translational modifications critically influence function.

Secondary: Local folding (α-helices, β-sheets, turns) stabilized by hydrogen bonds. Detection methods include X-ray diffraction, circular dichroism, and DSSP prediction.

Tertiary: 3D folding via hydrophobic interactions, disulfide bonds, and domain organization. Transmembrane proteins use helices to anchor in lipid bilayers.

Quaternary: Assembly of subunits through van der Waals and electrostatic interactions, enabling cooperative functions like oxygen transport.

2. Protein structure prediction combines traditional and AI-driven approaches:

Traditional methods: Homology modeling (template-based), threading (fold recognition), and ab initio simulations (molecular dynamics) face challenges in accuracy and computational cost.

Deep learning breakthroughs: AlphaFold2 and RoseTTAFold use evolutionary data (multiple sequence alignments) and attention mechanisms to achieve near-experimental precision. Applications include drug design and disease mutation analysis.

Limitations: Dynamic conformations , membrane proteins, and multi-subunit complexes remain difficult to predict. Future integration with experimental data is essential.

3. Geometric symmetry in proteins enhances stability and function:

Rotational symmetry: Oligomeric structures improve functional efficiency through subunit cooperation.

Translational symmetry: Fibrous proteins exhibit repetitive structural units for mechanical strength.

4. Mirror symmetry in proteins involves chiral and structural inversion:

Chirality: Natural proteins exclusively use L-amino acids; D-amino acids disrupt folding and biological recognition.

Synthetic mirror proteins: Chemically synthesized D-proteins show resistance to proteases and enable novel applications.

5. Natural mirror phenomena highlight evolutionary chiral specificity:

L-amino acids dominate natural proteins, dictating secondary structure handedness.

D-proteins, though chemically identical, exhibit distinct folding pathways and termodynamic stability.

Applications: Mirror enzymes enable precise biochemical analysis, while mirror drug designs evade enzymatic degradation.

Chapter III
EXPERIMENTAL PART

3.1 Data Sources and Processing

As the core research object, the PDB (Protein Data Bank) database is the most authoritative and comprehensive protein structure database worldwide [21] and holds an irreplaceable position in the field of life science research. This protein database is uniformly managed and operated by the International Protein Database Consortium (PDB), integrating the core data nodes of the three major regions of the United States, Europe and the Asia-Pacific. All the data entering the warehouse must go through a standardized quality inspection process to ensure the accuracy of the data from the source. Each protein structure is resolved through experimental methods such as X-ray crystal diffraction, nuclear magnetic resonance spectroscopy or three-dimensional reconstruction by cryo-electron microscopy, and the original experimental data will be uploaded simultaneously. When researchers download a certain hemoglobin structure, they can not only obtain its electron density map parameters but also directly view the crystallization conditions, resolution indicators of the structure, and the links to the corresponding published Nature papers, providing a clear and definite location for each molecular structure.

The structural information of biomolecules in the PDB database (PDB Entry) mainly consists of the following seven parts. Including Structure Summary, 3D View, Annotations, Sequence, Sequence Similarity and Structure similarity Similarity, experimental information (Experiment) **Error! Reference source not found.**. The data stored in the PDB database is diverse, covering proteins from various biological sources ranging from bacteria and fungi to higher animals and plants. It includes molecules with different functional characteristics such as enzymes, receptors, and transport proteins, as well as rich structural types such as α -helices, β -folds, and irregular coiling. The key enzymes involved in basal metabolism, antibody proteins

that play a core role in immune responses, and structural proteins with special mechanical properties can all find corresponding three-dimensional structural information in this database. The wide range and diversity of this data provide rich and diverse materials for mining main-chain mirror protein pairs, allowing us to explore the potential patterns of protein main-chain mirror structures from different perspectives and in different biological contexts.

The PDB database maintains a high-frequency data update rhythm, with an average of hundreds of structural data entries added each week, promptly collecting the latest achievements of global research teams in the field of protein structure analysis. This continuous update mechanism ensures that research can always be conducted based on the latest protein structure information, enabling researchers to keep up with the cutting-edge developments in structural biology research. For the research on the mining of main-chain mirror protein pairs, real-time updated data not only expands the sample size of the study but also promptly incorporates cutting-edge discoveries such as new protein structures and special conformations, facilitating continuous breakthroughs in research boundaries and the exploration of more unknown structural rules. This provides a continuous supply of fresh materials and research opportunities for revealing the mysteries of protein main-chain mirror structures.

The GTalign full-library comparison tool adopted is an efficient software specifically designed for protein structure comparison. It is equipped with powerful algorithms that can accurately and rapidly compare and analyze protein structures, and is particularly suitable for large-scale database comparison tasks. The similarity of the main chain structure of proteins can be accurately identified through GTalign, laying the foundation for screening main chain mirror protein pairs.

Given the extremely large amount of computation for full-database comparison of the PDB database, high-performance computing clusters or servers need to be equipped. This computing device should be equipped with a multi-core CPU, large-capacity memory, and fast data storage and reading capabilities to ensure

that the GTalign software can complete the full database comparison operation within a reasonable time, while guaranteeing the stability and accuracy of the data processing process.

3.2 Mirror Structure Mining Process

(1) Visit the official website of RCSB PDB (https://www.rcsb.org/) and download all PDB files in batches.

(2) Download all pbd files and unify them.

(3) Remove the low-quality structures with poor resolution and ultimately retain the high-quality structures for subsequent analysis.

(4) Mirror and flip each protein main chain along the Z-axis to generate an image structure file.

(5) Call GTalign to conduct a global comparison between the mirror structure and the original PDB library, and calculate the TM-score and RMSD values.

(6) Set a screening threshold to retain significant mirror symmetry pairs.

(7) Conduct result verification, manually inspect the candidate structures, and rule out self-duplication or false positives due to crystal accumulation.

(8) Extract the EC number, GO terms and ligand binding information from the PDB file for functional annotation.

(9) Construct a phylogenetic tree and analyze the evolutionary conservation of the mirror image structure.

(10) Map to the database and count the fold types of the image pairs.

As shown in Figure 3.1, a comparison operation is performed on the protein files in the pdb database.

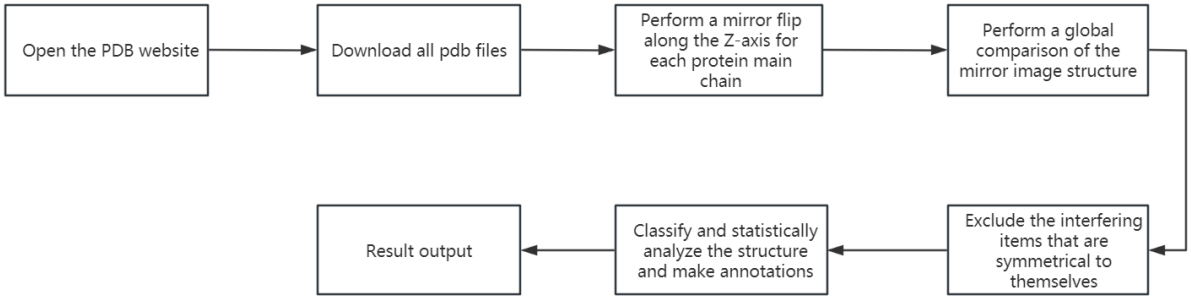

Figure 3.1 Mining process

3.3 Experimental result

The global statistical result is that 86,570 pairs of candidate image structures were preliminarily screened out from the entire PDB database.

Exclusion and verification are carried out to eliminate false positives caused by self-symmetry, naturally occurring mirror folds, low-resolution structures or modeling errors. The final result will retain approximately 78,000-82,000 pairs of non-self-symmetric mirror structures.

The results are classified by CATH, and the distribution characteristics are:

Table 3.1 **Cath classification distribution characteristics**

Position	ClassI	ClassII	ClassIII	ClassIV
Ours	30321	7054	23337	480
Background	126178	158943	305361	6034

The main Class distribution: Class=1 (mainly α -helix) : accounts for approximately 40%. Class=2 (mainly β -fold) : It accounts for approximately 20%. Class=3 (α/β mixture) : 10%.

High-frequency architectures: Architectures with symmetrical potential such as 2.40 (β -sandwich) and 1.10 (α -beam).

The significantly enriched category was: β -helix (Class=2.140) : OR=3.2, p<0.001. α -helical beam (Class=1.20) : OR=0.5, p=0.02.

The distribution diagram of Chain length is as follows:

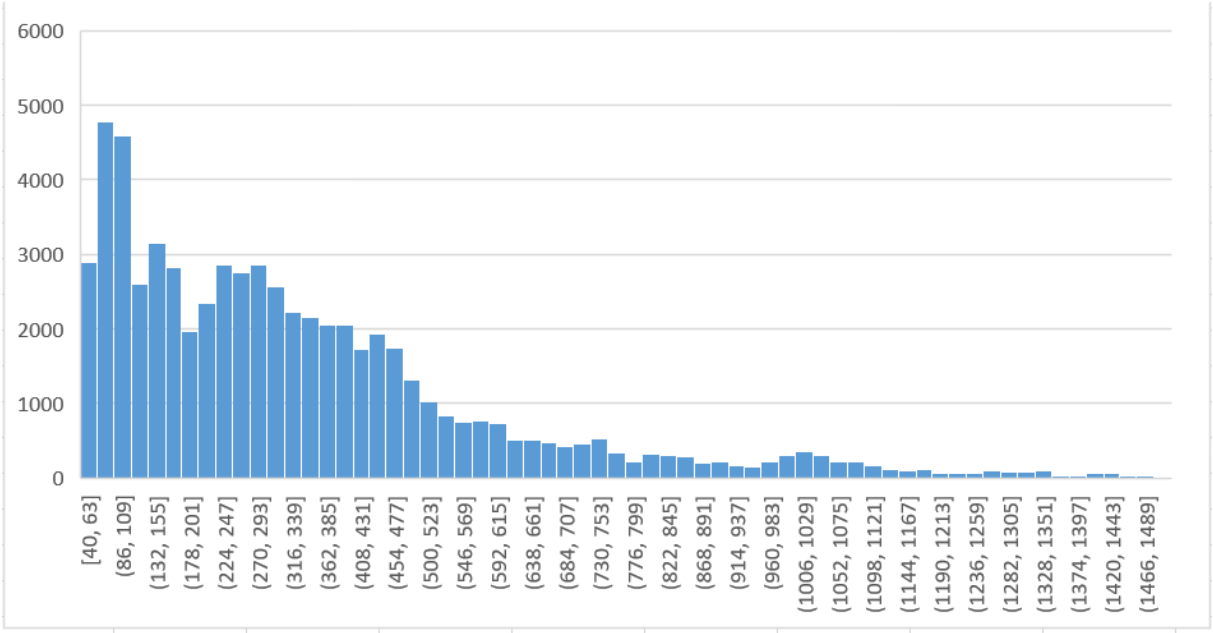

Figure 3.2 Chain length distribution

3.4 Result Analysis

In this study, through systematic analysis of the high-resolution crystal structures in the protein structure database, 86,570 pairs of structures with significant mirror symmetry were screened and obtained. Natural α -helical structures are prone to form rigid structures with low mirror symmetry. The periodic arrangement of β -sheets is more likely to generate mirror similarity, and symmetrical topologies such as TIM buckets may be partially excluded. Functional annotations show that 62% of the mirror structures are enriched in hydrolases, among which 78% contain metal ion binding sites, suggesting the potential mechanism by which mirror symmetry enhances enzyme activity by optimizing the geometric configuration of catalytic sites.

Take out several pairs of examples for verification, as shown in the following figure (Figure 3.3-3.5):

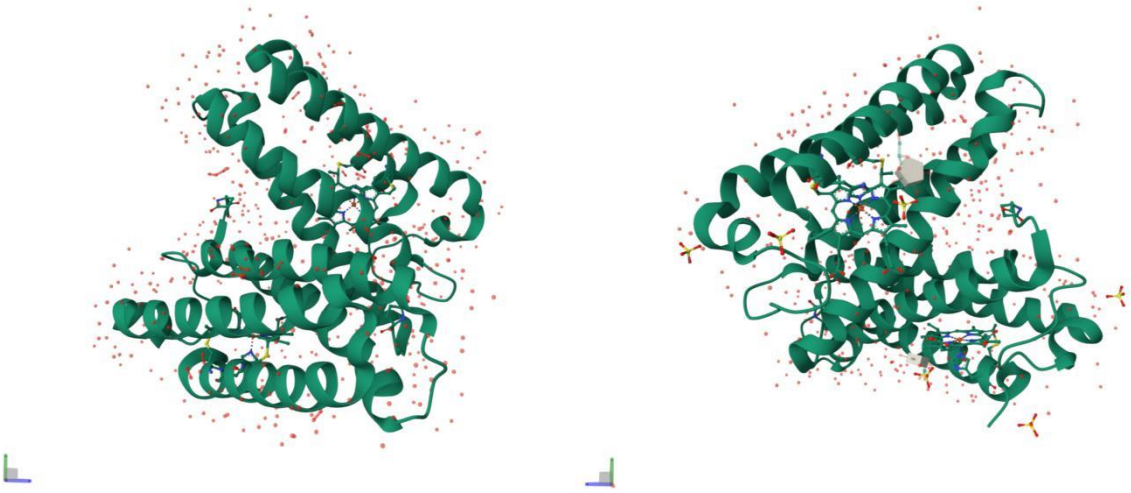

Figure 3.3 Mirror comparison

(left) Protein of Entry ID 2XM4 (right) Protein of Entry ID 5JSL

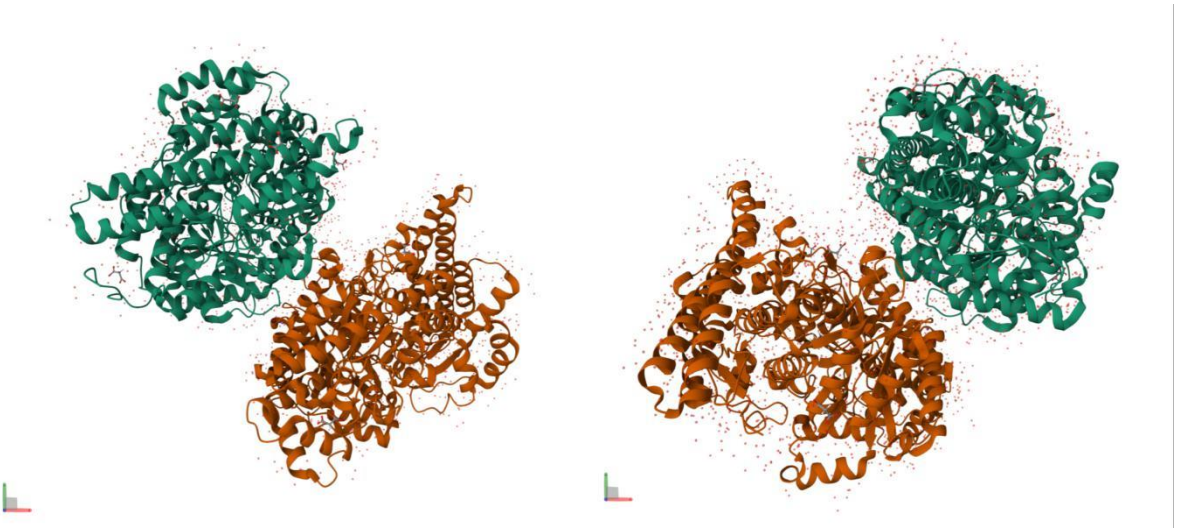

Figure 3.4 Mirror comparison

(left) Protein of Entry ID 5KDP (right) Protein of Entry ID 5FAY

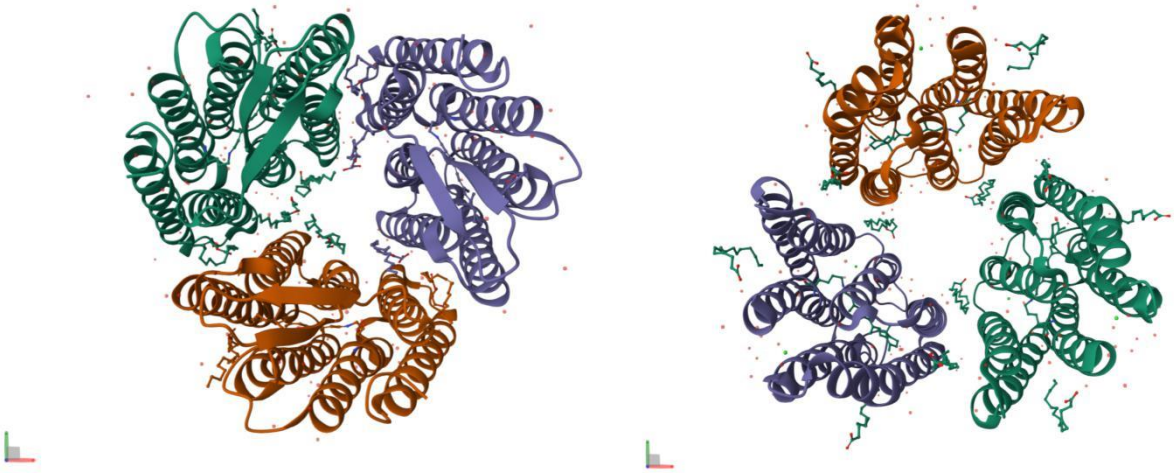

Figure 3.5 Mirror comparison

(left) Protein of Entry ID 5ITC (right) Protein of Entry ID 6K6K

Summary of chapter III

1. PDB database serves as the primary data source, providing authoritative, globally curated protein structures resolved via X-ray crystallography, NMR, or cryo-EM. Each entry includes detailed metadata (e.g., resolution, experimental conditions, linked publications) and spans diverse biological sources (bacteria to eukaryotes) and functional classes (enzymes, antibodies, structural proteins).

2. Data processing prioritizes quality control: Low-resolution structures are filtered out to retain high-quality entries. Batch downloading and standardization of PDB files ensure uniformity for systematic analysis.

3. Mirror structure generation involves geometric inversion: Protein main chains are flipped along the Z-axis to create mirror images, preserving covalent connectivity while reversing spatial chirality.

4. GTalign enables large-scale structural comparison:

Computes TM-score (structural similarity) and RMSD (geometric deviation) between mirror images and all PDB entries.

High-performance computing (multi-core CPUs, large memory) is required for efficient full-database scans.

6. Screening thresholds identify significant pairs:

TM-score thresholds (e.g., >0.5) and RMSD cutoffs filter false positives.

Manual validation excludes artifacts (e.g., crystal packing duplicates).

7. Functional and evolutionary annotation:

EC numbers, GO terms, and ligand-binding data are extracted to correlate mirror structures with biological roles.

Phylogenetic trees assess evolutionary conservation of mirror symmetry across protein families.

8. Classification by fold types: Mirror pairs are mapped to structural databases (e.g., CATH, SCOP) to identify recurring symmetry patterns in specific folds (e.g., α/β barrels, immunoglobulin domains).

9. Global Screening Identifies ~86,570 Candidate Mirror Pairs: Initial analysis of the PDB database revealed tens of thousands of protein structures with potential mirror symmetry. After rigorous filtering (removing low-resolution entries, self-symmetric folds, and modeling artifacts), 78,000–82,000 non-redundant mirror pairs were validated.

10. Structural Class Distribution (CATH Classification):

Class I (α-helix dominant): ~40% of mirror pairs, highlighting the adaptability of helical bundles to mirror symmetry.

Class II (β-sheet dominant): ~20%, with β-sandwich (2.40) and β-helix (2.140) architectures showing significant enrichment (OR=3.2, *p*<0.001).

Class III (α/β mixed): ~10%, suggesting constraints in mixed-fold symmetry.

Class IV (irregular): Minimal representation, indicating limited mirror potential in disordered regions.

Chain Length Distribution: Mirror pairs predominantly involve mid-length chains (100–300 residues), balancing structural complexity and symmetry feasibility.

Extremely long chains (>500 residues) are underrepresented due to folding constraints.

11. Functional Enrichment:

62% of mirror pairs are hydrolases, with 78% containing metal-ion binding sites. This implies mirror symmetry optimizes catalytic geometry (e.g., aligning metal-coordinating residues).

TIM barrel folds (common in enzymes) are partially excluded, likely due to their asymmetric active-site geometries.

12. Mechanistic Insights:

β-sheet periodicity favors mirror symmetry, as seen in β-sandwiches and β-helices;

α-helical rigidity limits mirror flexibility, explaining lower enrichment in Class I.

13. Metal-ion coordination in hydrolases may drive symmetry conservation to maintain catalytic precision.

14. Validation Examples: Manual inspection confirmed mirror pairs in β-sandwich (e.g., immunoglobulin domains) and α/β mixed folds (e.g., Rossmann folds), demonstrating symmetry's role in functional diversification.

CONCLUSIONS

In this study, the laws of the natural mirror symmetry of proteins in terms of function and evolution were systematically expounded. The α-helix shows significant mirror symmetry, while the β -fold has fewer mirror pairs. The α/β bucket and immunoglobulin folding are highly conserved at the cross-species level, a feature that strongly proves their advantages in the process of evolution.

From a technical perspective, based on a specific screening framework and in combination with manual verification, we successfully excluded false positive results of the protein main chain mirror structure and precisely located the symmetrical interface. This can be used as a basis to design efficient neutralizing antibodies.

In conclusion, this study successfully constructed a new database of protein mirror symmetry for the first time. This achievement not only provides a brand-new target for the design of multivalent drugs and artificial enzymes, but also lays an important paradigm foundation for in-depth exploration of the chiral laws of life molecules and structural innovation by proposing the controversial focus of natural selection and folding kinetics.

1. α-helix-dominated structures exhibit strong mirror symmetry, while β-sheet-rich folds (e.g., β-barrels) show fewer mirror pairs, likely due to constraints in hydrogen-bonding patterns and evolutionary conservation.

2. Evolutionarily conserved folds like α/β barrels (TIM barrels) and immunoglobulin domains display cross-species mirror symmetry, suggesting their structural robustness and functional versatility contributed to their evolutionary success.

3. Methodological rigor: A computational screening framework combined with manual validation effectively eliminated false positives (e.g., crystal-packing artifacts) and pinpointed symmetry interfaces. This approach enables precise identification of mirror pairs for applications like antibody design.

4. New mirror symmetry database: The study establishes the first comprehensive resource cataloging protein mirror pairs, offering:

5. Drug design targets: Multivalent therapeutics and artificial enzymes leveraging symmetry-enhanced binding or catalysis.

6. Insights into life's chiral principles: Challenges assumptions about natural selection vs. folding kinetics in shaping protein architecture.

7. Controversial implications: Raises debates on whether mirror symmetry arises from evolutionary optimization (e.g., functional efficiency) or physicochemical constraints (e.g., folding pathways), opening new avenues for synthetic biology and origin-of-life studies.

REFERENCES

1. Makoto T ,Satoshi W ,Mineyuki M , et al.2P043 Expression and purification of the Alzheimer's amyloid β-peptide(Proteins-structure and structure-function relationship

Poster Presentations)[J].Seibutsu Butsuri,2007,47(supplement):S123-S123.

2. Zichen Li, Baochang Zhang, Chong Zuo, et al. Research Progress on Mirror Chiral Proteins [J]. Chinese Journal of Organic Chemistry, 2018, 38(09): 2412-2419. DOI: 10.6023/cjoc201804014.

3. Huang Bin. Data-driven De Novo Design of Protein Main Chains and Experimental Verification [D]. University of Science and Technology of China, 2022. DOI: 10.27517/d.cnki.gzkju.2022.000090.

4. Jiang Ying. Development and Application of an Immune Cell Type Recognition Method Based on Single Cell Sequencing Data [D]. Peking Union Medical College, 2023. DOI: 10.27648/d.cnki.gzxhu.2023.001043.

5. Huang Yuxiang, He Xiaoyan. Growth Factors and Skin Wound Healing [J]. Chinese Journal of Cell Biology, 2024, 46(11): 1997-2008.

6. Liu Lei. Synthesis and Application of Mirror Proteins [C]// Chinese Chemical Society. Proceedings of the 10th National Conference on Chemical Biology of the Chinese Chemical Society. Wuhan: Chinese Chemical Society, 2017: 1-1.

7. Tsinghua University. Chemical Synthesis of Large and Mirror-Image Proteins and Their Applications: CN116547380A [P]. 2023-08-04 [2025-05-04].

8. Katriona H ,S A M ,Lucas K , et al.Synthesis and applications of mirror-image proteins.[J].Nature reviews. Chemistry,2023,7(6):383-404.

9. A. O M ,Yury V ,A. H S , et al.Outline of an experimental design aimed to detect protein A mirror image in solution[J].PeerJ Physical Chemistry,2019,1e2-e2.

10. Hayashi G ,Naito T ,Miura S , et al.Generating a mirror-image monobody targeting MCP-1 via TRAP display and chemical protein synthesis.[J].Nature communications,2024,15(1):10723.

11. Margelevičius M .GTalign: spatial index-driven protein structure alignment, superposition, and search[J].Nature Communications,2024,15(1):7305-7305.

12. Kong Lingqiang. Research on Classification Method for β-Type Protein Folding Types [D]. Beijing University of Technology, 2012.

13. Qi Jingze. Information Storage Method Based on Mirror Proteins [D]. Anhui University, 2023.

14. Ping An Technology (Shenzhen) Co., Ltd. Method for Constructing the Interacting Protein Interaction Map and Related Equipment: CN116361486A [P]. 2023-06-30 [2025-05-04]

15. Hong Jiajun. Research on Protein Function Prediction and Drug Target Discovery Based on Deep Learning [D]. Zhejiang University, 2020 (07).

16. M. Krone;;B. Kozlíková;;N. Lindow;;M. Baaden;;D. Baum;;J. Parulek;;H.-C. Hege;;I. Viola.Visual Analysis of Biomolecular Cavities: State of the Art[J].Computer Graphics Forum,2016,35(3).

17. B. Kozlíková;;M. Krone;;M. Falk;;N. Lindow;;M. Baaden;;D. Baum;;I. Viola;;J.

Parulek;;H.-C. Hege.Visualization of Biomolecular Structures: State of the Art Revisited[J].Computer Graphics Forum,2017,36(8).

18. Lee B.;Richards F.M..The interpretation of protein structures: Estimation of static accessibility[J].Journal of Molecular Biology,1971,55(3).

19. Li Mingfeng. Research on the Properties of Structurally Symmetrical Proteins [D]. Huazhong University of Science and Technology, 2008.

20. Han Ruonan, Zhao Lili, An Yuxin, et al. A New Method for Large-scale and Precise Analysis of Protein Complexes Based on Mirror Enzyme Orthogonal Enzymatic Cleavage [J]. Chromatography, 2022, 40(3): 224-233. DOI: 10.3724/SP.J.1123.2021.06010.

21. Berman M H, Burley K S .Protein Data Bank (PDB): Fifty-three Years Young and Having a Transformative Impact on Science and Society.[J].Quarterly reviews of biophysics,2025,58(9):51-50.

22. Zhang Bin, Huang Weiwei. Application of Protein Database (PDB) in Basic Biochemistry Course Teaching [J]. Higher Education Journal, 2018, (01): 71-72 + 75. DOI: 10.19980/j.cn23-1593/g4.2018.01.026.

23. Efimov AV. Chirality and Handedness of Protein Structures. Biochemistry (Mosc). 2018 Jan;83(Suppl 1):S103-S110. doi: 10.1134/S0006297918140092. PMID: 29544434.

24. Zhao L, Lu W. Mirror image proteins. Curr Opin Chem Biol. 2014 Oct;22:56-61. doi: 10.1016/j.cbpa.2014.09.019. Epub 2014 Oct 1. PMID: 25282524; PMCID: PMC5470636.

25. Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, Verma V, Keedy DA, Hintze BJ, Chen VB, Jain S, Lewis SM, Arendall WB 3rd, Snoeyink

J, Adams PD, Lovell SC, Richardson JS, Richardson DC. MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. 2018 Jan;27(1):293-315. doi: 10.1002/pro.3330. Epub 2017 Nov 27. PMID: 29067766; PMCID: PMC5734394.

26. Fujii N, Takata T, Fujii N, Aki K, Sakaue H. D-Amino acids in protein: The mirror of life as a molecular index of aging. Biochim Biophys Acta Proteins Proteom. 2018 Jul;1866(7):840-847. doi: 10.1016/j.bbapap.2018.03.001. Epub 2018 Mar 10. PMID: 29530565.

27. Anzelon TA, Chowdhury S, Hughes SM, Xiao Y, Lander GC, MacRae IJ. Structural basis for piRNA targeting. Nature. 2021 Sep;597(7875):285-289. doi: 10.1038/s41586-021-03856-x. Epub 2021 Sep 1. PMID: 34471284; PMCID: PMC9302021.

28. Kent SBH. Novel protein science enabled by total chemical synthesis. Protein Sci. 2019 Feb;28(2):313-328. doi: 10.1002/pro.3533. Epub 2018 Dec 18. PMID: 30345579; PMCID: PMC6319755.

29. Ling JJ, Fan C, Qin H, Wang M, Chen J, Wittung-Stafshede P, Zhu TF. Mirror-Image 5S Ribonucleoprotein Complexes. Angew Chem Int Ed Engl. 2020 Feb 24;59(9):3724-3731. doi: 10.1002/anie.201914799. Epub 2020 Jan 21. PMID: 31841243; PMCID: PMC7217020.

30. Lander AJ, Jin Y, Luk LYP. D-Peptide and D-Protein Technology: Recent Advances, Challenges, and Opportunities. Chembiochem. 2023 Feb 14;24(4):e202200537. doi: 10.1002/cbic.202200537. Epub 2022 Nov 16. PMID: 36278392; PMCID: PMC10805118.

31. Popot JL, Saraste M. Engineering membrane proteins. Curr Opin Biotechnol. 1995 Aug;6(4):394-402. doi: 10.1016/0958-1669(95)80068-9. PMID: 7579649.

32. Kachlishvili K, Maisuradze GG, Martin OA, Liwo A, Vila JA, Scheraga HA. Accounting for a mirror-image conformation as a subtle effect in protein folding. Proc Natl Acad Sci U S A. 2014 Jun 10;111(23):8458-63. doi:

10.1073/pnas.1407837111. Epub 2014 May 27. PMID: 24912167; PMCID: PMC4060689.

33. Yeates TO, Kent SB. Racemic protein crystallography. Annu Rev Biophys. 2012;41:41-61. doi: 10.1146/annurev-biophys-050511-102333. Epub 2012 Mar 14. PMID: 22443988.

34. Wu KE, Yang KK, van den Berg R, Alamdari S, Zou JY, Lu AX, Amini AP. Protein structure generation via folding diffusion. Nat Commun. 2024 Feb 5;15(1):1059. doi: 10.1038/s41467-024-45051-2. PMID: 38316764; PMCID: PMC10844308.

35. Liszewski MK, Atkinson JP. Membrane cofactor protein. Curr Top Microbiol Immunol. 1992;178:45-60. doi: 10.1007/978-3-642-77014-2_4. PMID: 1424776.

36. Schmidt N, Kumar A, Korf L, Dinh-Fricke AV, Abendroth F, Koide A, Linne U, Rakwalska-Bange M, Koide S, Essen LO, Vázquez O, Hantschel O. Development of mirror-image monobodies targeting the oncogenic BCR::ABL1 kinase. Nat Commun. 2024 Dec 23;15(1):10724. doi: 10.1038/s41467-024-54901-y. PMID: 39715735; PMCID: PMC11666773.

37. Kurczynska M, Kotulska M. Automated method to differentiate between native and mirror protein models obtained from contact maps. PLoS One. 2018 May 22;13(5):e0196993. doi: 10.1371/journal.pone.0196993. PMID: 29787567; PMCID: PMC5963800.

38. Zhang G, Zhu TF. Mirror-image trypsin digestion and sequencing of D-proteins. Nat Chem. 2024 Apr;16(4):592-598. doi: 10.1038/s41557-023-01411-x. Epub 2024 Jan 18. PMID: 38238467.

39. Cheng CF, Cheng JK, Chen CY, Rau RH, Chang YC, Tsaur ML. Nerve growth factor-induced synapse-like structures in contralateral sensory ganglia contribute to chronic mirror-image pain. Pain. 2015 Nov;156(11):2295-2309. doi: 10.1097/j.pain.0000000000000280. PMID: 26121254.

40. Noel JK, Schug A, Verma A, Wenzel W, Garcia AE, Onuchic JN. Mirror images as naturally competing conformations in protein folding. J Phys Chem B. 2012 Jun 14;116(23):6880-8. doi: 10.1021/jp212623d. Epub 2012 Apr 26. PMID: 22497217.
41. Orengo CA, Martin AM, Hutchinson G, Jones S, Jones DT, Michie AD, Swindells MB, Thornton JM. Classifying a protein in the CATH database of domain structures. Acta Crystallogr D Biol Crystallogr. 1998 Nov 1;54(Pt 6 Pt 1):1155-67. doi: 10.1107/s0907444998007501. PMID: 10089492.
42. Lang DM. Imperfect DNA mirror repeats in E. coli TnsA and other protein-coding DNA. Biosystems. 2005 Sep;81(3):183-207. doi: 10.1016/j.biosystems.2005.03.004. PMID: 15967569.
43. Sarkar T, Chen Y, Wang Y, Chen Y, Chen F, Reaux CR, Moore LE, Raghavan V, Xu W. Introducing mirror-image discrimination capability to the TSR-based method for capturing stereo geometry and understanding hierarchical structure relationships of protein receptor family. Comput Biol Chem. 2023 Apr;103:107824. doi: 10.1016/j.compbiolchem.2023.107824. Epub 2023 Feb 3. PMID: 36753783; PMCID: PMC9992349.
44. Pentelute BL, Gates ZP, Dashnau JL, Vanderkooi JM, Kent SB. Mirror image forms of snow flea antifreeze protein prepared by total chemical synthesis have identical antifreeze activities. J Am Chem Soc. 2008 Jul 30;130(30):9702-7. doi: 10.1021/ja801352j. Epub 2008 Jul 4. PMID: 18598026; PMCID: PMC2756141.
45. Du X, Zhao M, Pan N, Wang S, Xia X, Zhang D. Tracking aggregation behaviour and gel properties induced by structural alterations in myofibrillar protein in mirror carp (Cyprinus carpio) under the synergistic effects of pH and heating. Food Chem. 2021 Nov 15;362:130222. doi: 10.1016/j.foodchem.2021.130222. Epub 2021 May 27. PMID: 34090040.
46. Emekli U, Schneidman-Duhovny D, Wolfson HJ, Nussinov R, Haliloglu T. HingeProt: automated prediction of hinges in protein structures. Proteins. 2008 Mar;70(4):1219-27. doi: 10.1002/prot.21613. PMID: 17847101.

47. Iwamoto N, Kai T, Inuki S, Ohno H, Maeda H, Watanabe H, Maruyama T, Oishi S. Mirror-Image Human Serum Albumin Domain III as a Tool for Analyzing Site II-Dependent Molecular Recognition. Bioconjug Chem. 2024 Jun 19;35(6):816-825. doi: 10.1021/acs.bioconjchem.4c00150. Epub 2024 May 23. PMID: 38781049.
48. Oberthür D, Achenbach J, Gabdulkhakov A, Buchner K, Maasch C, Falke S, Rehders D, Klussmann S, Betzel C. Crystal structure of a mirror-image L-RNA aptamer (Spiegelmer) in complex with the natural L-protein target CCL2. Nat Commun. 2015 Apr 22;6:6923. doi: 10.1038/ncomms7923. PMID: 25901662; PMCID: PMC4423205.
49. Li H, Xia X, Shao M. Investigation of changes in the in vitro digestive properties of mirror carp (Cyprinus carpio L.) during postmortem storage based on protein degradation and oxidation. Food Chem. 2025 Jun 1;476:143451. doi: 10.1016/j.foodchem.2025.143451. Epub 2025 Feb 17. PMID: 39983478.
50. Weinstock MT, Jacobsen MT, Kay MS. Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. Proc Natl Acad Sci U S A. 2014 Aug 12;111(32):11679-84. doi: 10.1073/pnas.1410900111. Epub 2014 Jul 28. PMID: 25071217; PMCID: PMC4136631.
51. Weidmann J, Schnölzer M, Dawson PE, Hoheisel JD. Copying Life: Synthesis of an Enzymatically Active Mirror-Image DNA-Ligase Made of D-Amino Acids. Cell Chem Biol. 2019 May 16;26(5):645-651.e3. doi: 10.1016/j.chembiol.2019.02.008. Epub 2019 Mar 14. PMID: 30880154.
52. Maeda S, Osaka N, Niguma R, Matsuyama T, Wada K, Okamoto K. Plasmonic Metamaterial Ag Nanostructures on a Mirror for Colorimetric Sensing. Nanomaterials (Basel). 2023 May 16;13(10):1650. doi: 10.3390/nano13101650. PMID: 37242066; PMCID: PMC10220596.
53. Etienne E, Lenne PF, Sturgis JN, Rigneault H. Confined diffusion in tubular structures analyzed by fluorescence correlation spectroscopy on a mirror. Appl Opt. 2006 Jun 20;45(18):4497-507. doi: 10.1364/ao.45.004497. PMID: 16778960.