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 **Theoretical Research on Substrate Specificity Differences within the Feruloyl Esterase Family**

First (Bachelor's) level of higher education
Specialty 162 "Biotechnology and Bioengineering"
Educational and professional program "Biotechnology"

Completed: student of group BEBT-20
Li WEILONG

Scientific supervisor
Iryna VOLOSHYNA, Ph.D., As. prof.

Reviewer
Liubov ZELENA, Ph.D., As. prof.

Kyiv 2024

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 Department of Biotechnology, Leather and Fur, Professor,
Doctor of Technical Science
Olena MOKROUSOVA
_____
«___»_____2024

ASSIGNMENTS
FOR THE QUALIFICATION THESIS
Li Weilong

1. Thesis topic **Theoretical Research on Substrate Specificity Differences within the Feruloyl Esterase Family**
Scientific supervisor Iryna Voloshyna, Ph.D., As. prof.

approved by the order of KNUTD "_____" _____2024, №_____

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_____

EXECUTION SCHEDULE

№	The name of the stages of the qualification thesis	Terms of performance of stage	Note on performance
1	Introduction	From 01 April 2024 to 11 April 2024	
2	Chapter 1. Literature review	From 06 April 2024 to 20 April 2024	
3	Chapter 2. Object, purpose, and methods of the study	From 21 April 2024 to 30 April 2024	
4	Chapter 3. Experimental part	From 01 May 2024 to 10 May 2024	
5	Conclusions	From 07 May 2024 to 12 May 2024	
6	Draw up a bachelor's thesis (final version)	From 20 May 2024 to 25 May 2024	
7	Submission of qualification work to the supervisor for feedback	From 25 May 2024 to 27 May 2024	
8	Submission of bachelor's thesis to the department for review (14 days before the defense)	27 May 2024	
9	Checking the bachelor's thesis for signs of plagiarism (10 days before the defense)	31 May 2024	
10	Submission of bachelor's thesis for approval by the head of the department (from 7 days before the defense)	03 June 2024	

I am familiar with the task:

Student _____ Li WEILONG

Scientific supervisor _____ Iryna VOLOSHYNA

SUMMARY

Weilong Li. Theoretical Research on Substrate Specificity Differences within the Feruloyl Esterase Family – Manuscript.

Qualification thesis on the specialty 162 «Biotechnology and Bioengineering». – Kyiv National University of Technologies and Design, Kyiv, 2024.

This study comprehensively investigated the substrate specificity differences among four feruloyl esterases using computational biology methods, offering novel insights into enzymatic functions, structural factors influencing substrate specificity, and the optimization of enzyme performance.

The research aimed to elucidate the molecular mechanisms underlying the substrate specificity of feruloyl esterases (FAEs) and to assess the impact of disulfide bonds on their activity. FAEs are crucial in bioenergy and biocatalysis, yet the understanding of their substrate specificity remains limited. The study comprised three parts. First, the analysis of substrate specificity differences among four FAEs; second, the evaluation of the effects of disrupting disulfide bonds on FAE activity; and finally, the proposal of computational strategies for improving FAE applications.

By employing homology modeling, active site analysis, molecular docking, and dynamic simulations, this study constructed FAE structural models and simulated FAE-substrate interactions. The results revealed that specific amino acid residues play a key role in substrate recognition, and the disruption of disulfide bonds can affect the enzyme's substrate affinity.

The discussion section explores the molecular mechanisms of FAEs' substrate specificity and the theoretical and practical contributions of computational predictions to enzyme engineering. The study suggests future research directions to deepen the understanding of FAE substrate specificity and offers new strategies for synthetic biology and biocatalysis.

In conclusion, this research provides new insights into the functions of FAEs, the structural basis of substrate specificity, and how to optimize enzyme performance through computational methods. performance through computational methods.

Keywords: Computational Biology, Feruloyl Esterase, Substrate Preference, Homology Modeling, Molecular Docking, Molecular Dynamics Simulation, Point Mutation, Enzyme Activity

TABLE OF CONTENTS

INTRODUCTION

This study employs computational biology methods to conduct theoretical research on the substrate specificity differences within the Feruloyl Esterase (FAE) family, a subclass of carboxylesterases that hydrolyze ester bonds between ferulic acid and polysaccharides in plant cell walls. The objective is to leverage computational analysis to understand the diversity and functional attributes of FAEs, enzymes with significant industrial applications in the food, feed, pharmaceutical, and biofuel sectors due to their ability to release phenolic compounds with antioxidant and anti-inflammatory properties. The innovative aspect of this study lies in its use of computational biology to shed light on FAEs' substrate specificity, potentially enhancing their industrial utility and efficiency. The goals and objectives of this research are twofold. First, we aim to optimize Feruloyl Esterases (FAEs) for a variety of applications by gaining a more comprehensive understanding of their genomic and functional characteristics. Second, we intend to investigate the impact of creating and knocking out disulfide bonds on the performance of ferulic esterase.

The relevance of the topic is theoretical studies on substrate-specific differences in the feruloyl esterase family using computational biology methods

The purpose of the study is to leverage computational biology to deepen the understanding of FAEs' diversity and functionality, enhancing their industrial applications.

The goals and objectives of the study are to conduct theoretical research on the substrate specificity differences within FAEs and understand their functional attributes through computational analysis. Also, the effect of disulfide bonds on enzymes was studied by adding or knocking down disulfide bonds.

The object of the study The Feruloyl Esterase (FAE) family of enzymes.

The subject of the study the substrate specificity differences within the FAE family.

Research methods the computational biology methods.

The scientific novelty the use of computational biology to elucidate the substrate specificity of FAEs, which could potentially enhance their efficiency and utility in various industries.

The practical significance of the results obtained is the potential optimization of FAEs for improved use in the food, feed, pharmaceutical, and biofuel sectors through a better understanding of their genomic and functional characteristics.

CHAPTER 1
LITERATURE REVIEW

1.1 Introduction to Feruloyl Esterase

Ferulic acid esterases (FAEs), also known as ferulic acid esterases, are a subclass of carboxylesterases capable of hydrolyzing ester bonds between ferulic acid and polysaccharides in plant cell walls. They belong to a larger class of enzymes known as hydrolases[1].

With the widespread use of biotechnology in industry, agriculture and medicine, there is a growing demand for research and development of biological enzymes as catalysts. Feruloyl esterase (FAE) is a class of enzymes that play a key role in the degradation of plant cell walls by hydrolyzing the ester bonds between phenolic acids and polysaccharides, releasing phenolic compounds with antioxidant and anti-inflammatory effects, such as ferulic acid (FA) [2], and FAEs have a wide range of applications in food, feed, pharmaceutical and biofuel industries with a wide range of applications[3]. Feruloyl Esterases (FAEs, EC 3.1.1.73) belong to a subclass of carboxylesterases and show their unique catalytic role in the extraction of ferulic acid (FA) and its derivatives from plant biomass, paper making, biofuel production and food industry[4]. These enzymes effectively hydrolyze hydroxycinnamate esters in plant cell walls to release commercially valuable aromatic compounds which have a wide range of applications in several fields[5].

1.1.1 Biological functions of ferulic acid esterases

Ferulic acid esterases (FAEs), also known as ferulic acid ester hydrolases, are important members of the esterase family. These enzymes play a crucial role in the degradation of plant cell walls by hydrolyzing the ferulic acid ester bonds connecting hemicellulose to lignin, pectin and other components. FAEs are widely found in a variety of organisms, including certain bacteria, fungi, and insects, which utilize these enzymes to break down plant material for energy. In the human gut microbiota, FAEs help break down dietary fiber and play an active role in maintaining gut health.

Industrially, FAEs are noted for their potential value in biofuel production, particularly their ability to degrade lignocellulosic biomass into fermentable sugars, a biomass that holds promise as a renewable resource to replace fossil fuels. FAEs also release ferulic acid with antioxidant properties from plant cell walls, a valuable compound with potential applications in food preservation and cosmetics.

1.1.2 Structure and function of ferulic acid esterases (FAEs)

Structurally, FAEs usually consist of a catalytic domain containing a conserved catalytic triad (serine, histidine and aspartic acid) responsible for ester hydrolysis. The active site of FAEs contains key residues involved in substrate binding and catalysis, including aromatic residues that interact with the phenolic rings of ferulic acid and polysaccharide substrates. The structural diversity of the FAE class results in their distinct substrate specificity and catalytic properties, highlighting the importance of understanding the molecular basis of substrate recognition by these enzymes.

Functionally, FAEs play a key role in plant cell wall degradation and biomass utilization. By breaking the ester bond between ferulic acid and polysaccharides, FAEs facilitate the release of valuable compounds, such as sugars and phenolic compounds, which can be further processed for various industrial applications. The ability of FAEs to alter plant biomass and enhance enzymatic degradation of lignocellulosic materials makes them valuable as biocatalysts in biorefinery processes.

1.2 Application of ferulic acid esterases

Ferulic acid esterases (FAEs) are a special class of enzymes that show great potential for industrial and medical applications by breaking down complex structures in plant cell walls, especially the ester bonds between ferulic acid and polysaccharides.

They play a key role in biofuel production, especially from lignocellulosic biomass. This biomass is abundant and renewable, but its complex structure makes

decomposition difficult. FAEs are able to degrade this complex structure into fermentable sugars, which can be converted into biofuels[6]. FAEs can also be used to improve the nutritional value of animal feed by breaking down plant material into a form that is easier for animals to digest[7]. FAEs also play an important role in the food and beverage industry, for example by releasing bound phenolic compounds from plant-based foods, enhancing their nutritional value and flavor profile, and in the maturation of alcoholic beverages such as wine and beer[8]. They also help in the removal of lignin, making fibers easier to separate and thus reducing chemical use and energy consumption[9].

1.3 Applications and Methods of Computational Biology in Enzymology Research

Computational biology, also known as bioinformatics, is a field that uses computational techniques to analyze, predict and interpret biological data. In enzymology research, computational biology plays a crucial role, not only in helping to reveal the structure, function and dynamic properties of enzymes, but also in designing and developing novel enzymes.

Its ability to predict the three-dimensional structure of an enzyme from its amino acid sequence allows researchers to visualize the active site and overall structure of the enzyme, which is critical for understanding enzyme function and substrate specificity. Molecular docking studies help identify key residues involved in substrate binding or catalysis by modeling the interaction of small molecules (e.g., substrates or inhibitors) with the enzyme, information that is important for designing more efficient enzymes or potent inhibitors. Computational methods can predict the function of an enzyme based on its sequence or structure, which is particularly useful for annotating newly discovered enzymes or enzymes with unknown functions.

1.3.1 Computational methods for homology modeling and structure prediction

Homology modeling, also known as comparative modeling, is based on the core principle that the structure of proteins is more similar to the protein sequence than the

protein sequence during the evolutionary process by searching for template proteins similar to the target protein sequence in protein databases through the method of sequence comparison. Once a suitable template is found, an initial model of the target protein can be constructed based on the results of the comparison between the target sequence and the template sequence, and the initial model can be optimized and refined by optimizing the energy function of the model, such as MODELLER or SWISS-MODEL, to obtain the final three-dimensional structural model[**Ошибка! Источник ссылки не найден.**].

In general, homology modeling results are more reliable when the identity of two sequences is greater than 30%[11], and even in the case of low sequence identity, homology modeling can still provide valuable structural information, especially in the absence of experimental structural data [12].

1.4.1 Computational methods for molecular docking and kinetic simulations

Molecular docking and kinetic simulation are two important tools in computational biology for exploring and studying molecular interactions.

Molecular docking is a computational process that focuses on predicting the optimal orientation of one molecule (usually a drug or substrate) when it binds to another molecule (usually a protein or enzyme) to form a stable complex. Molecular docking techniques have important applications in drug design and lead compound screening[13]. Molecular dynamics simulation is a computational method to study the dynamic properties of molecular systems by modeling the motions of atoms and molecules, which can provide information on conformational and energy changes of molecular systems under different conditions and help to understand the functional mechanisms of biomolecules[14].

These two tools provide valuable insights into molecular interactions at the atomic level, which are usually unavailable from experimental methods. With these tools, the combination of molecular docking and molecular dynamics simulations allows for a more comprehensive study of the ligand-receptor binding process and

binding mechanisms[15], providing theoretical support for drug design and enzymology studies.

1.4 Knowledge gaps on the molecular mechanisms of FAE substrate specificity

Although feruloyl esterases (FAEs) are of significant industrial importance, there remains a gap in knowledge regarding the molecular basis of their substrate specificity. The complex interplay between enzyme structure, substrate recognition, and catalytic activity poses challenges in elucidating the regulatory mechanisms of FAEs' substrate specificity. Furthermore, the structural diversity among different classes of FAEs further complicates the elucidation of determinants for specific substrate binding. Addressing these knowledge gaps is crucial for advancing the field of enzyme engineering and biocatalysis, and paves the way for designing FAE variants with tailored substrate specificity for diverse industrial applications.

1.5 Purpose, hypothesis and significance of the study

Ferulic acid esterases (FAEs) are a group of enzymes that play a key role in the degradation of complex plant materials, particularly by breaking the ester bond between hemicellulose and lignin. This activity is not only essential for the breakdown of natural organic matter, but also of great importance for industrial applications such as biofuel production, food processing and pharmaceuticals. Despite their importance, however, the molecular mechanisms underlying the substrate-specific preference of FAEs are still poorly understood, limiting a comprehensive understanding of their function and further application development, despite their great potential for application. This knowledge gap hinders our ability to effectively utilize these enzymes for industrial applications.

The aim of this study was to investigate the structural and functional differences among four classes of ferulic acid esterases (FAEs), A, B, C, and D, in order to elucidate the molecular basis of their substrate specificity. By comparing the amino acid sequences, three-dimensional structures and active site architectures of these FAE classes, the aim was to identify key residues and structural features that

contribute to their different substrate preferences. The comparative analysis seeks to unravel the structural determinants of substrate recognition and binding in different FAE classes, providing insights into the evolution of the substrate specificity of these enzymes, using computational biology methods to investigate the differences in substrate preference of the four specific FAEs and the effect of disulfide bonding around their active sites on their binding activity to model substrates. The aim is to elucidate the structural and functional features that lead to these differences. This involved generating detailed structural models of each enzyme, followed by molecular dynamics simulations and docking studies to predict and understand their interactions with various substrates.

FAEs typically have an α/β-hydrolase structure and often contain the Ser-Asp-His catalytic triad[16]. Their substrates are ester bonds linking phenolic acids or related cinnamic acids to plant cell wall polysaccharides[17]. The activity and substrate specificity of FAEs are influenced by their structural features, which can be analyzed by computational biology methods.

The differences in substrate specificity of A-, B-, C-, and D-type FAEs were explored by computational biology methods to reveal the molecular basis of their substrate-binding mechanisms, and by using molecular modeling, molecular dynamics simulations, and bioinformatics tools, we aimed to predict the effects of amino acid substitutions, and structural modifications on the substrate specificity and catalytic efficiency of FAEs.

Computational biology, as an emerging interdisciplinary science, provides new perspectives and tools for biological research through data analysis, mathematical modeling and computer simulation techniques. Homology modeling and molecular docking have been used to predict the three-dimensional structure and substrate binding sites of FAE[18]. These methods can help to understand the preference of FAE for different substrates, including the selectivity for phenolic acid esters with different substituents. In this work, computational biology methods such as homology modeling, molecular docking, and molecular dynamics simulations will be used to investigate the three-dimensional structures, active site characteristics, and interaction

patterns with substrates of FAEs of types A and B, C, and D. The results will be presented in the next section.

In addition to investigating the role of primary sequence and structural features in FAE substrate specificity, we aimed to assess the effect of disulfide bonds on enzyme activity and substrate specificity. Disulfide bonds play a key role in stabilizing protein structure and regulating enzyme function; therefore, disruption of these bonds may alter the catalytic properties and substrate preferences of FAEs. By selectively disrupting disulfide bonds in different classes of FAEs using mutagenesis, we aim to assess the effects of these modifications on enzyme activity, substrate binding affinity, and specificity. This experimental approach will provide valuable insights into the functional importance of disulfide bonds in FAEs and their potential role in shaping substrate specificity.

Through these studies, it is expected to reveal the molecular basis of the differences in substrate specificity of FAEs, provide a theoretical basis for the engineering modification and optimization of FAEs, and thus help in the development of FAEs applications in biomass energy conversion and biosynthesis.

The current status of domestic and international research shows that although the structure and function of FAEs have been studied to a certain extent, the in-depth understanding of their substrate specificity is still limited.

The aim of this study was to assess the effect of disulfide bond disruption on enzyme activity by investigating the structural factors underlying the substrate specificity of different FAE classes. Through an interdisciplinary approach integrating experimental and computational techniques, the molecular mechanisms of FAE substrate specificity are explored to facilitate the rational design of FAE variants with tailored catalytic properties.

CHAPTER 2

OBJECT, PURPOSE AND METHODS OF THE STUDY

2.1 Sequence search

In order to comprehensively collect and analyze the sequence information of different types of ferulic acid esterases, we searched and screened from several authoritative protein databases. First, we retrieved the structural information of the class A ferulic acid esterase AnFaeA from the RCSB Protein Data Bank (RCSB PDB, https://www.rcsb.org/) with the accession number 2BJH. The RCSB PDB is a database containing a large number of experimentally resolved three-dimensional structures of proteins, nucleic acids, and complexes, which provides a valuable resource for the study of structure-function relationships of proteins.

Next, we retrieved the sequence information of the class B ferulic acid esterase NcFae-I (accession number: Q9HGR3) and the class C ferulic acid esteraseFAEC_EMENI(accession number: Q5B2G3) from the UniProt Protein Data Bank (https://www.uniprot.org/) . UniProt is a comprehensive protein sequence and functional annotation database that integrates information from multiple sources to provide high-quality protein sequences and detailed functional annotations. By searching in UniProt, we obtained the complete sequence information of class B and class C ferulic acid esterases, as well as the related functional and evolutionary information.

Finally, we retrieved the sequence information of AcFAE, a class D ferulic acid esterase, from NCBI's Protein Data Bank (https://www.ncbi.nlm.nih.gov/) with the accession number XP_001274884.1. NCBI is a comprehensive bioinformatics resource platform, in which the protein database contains a large amount of protein sequence information from genome sequencing and annotation projects. By searching in NCBI, we obtained the complete sequence information of class D ferulic acid esterase, as well as the related genomic background and evolutionary information.

Table 2.1 – **Sequence search platform and information**

Name of enzyme	Search platform	Login number	Sequence of the enzyme
AnFaeA	RCSB PDB	2BJH	>2BJH_A_Feruloyl_Esterase_A ASTQGISEDLYNDLVEMATISQAAYADLCNIPSTIIKGE KIYNAQTDINGWILRDDTSKEIITVFRGTGSDT NLQLDTNYTLTPFDTLPQCNDCEVHGGYYIGWISVQD QVESLVKQQASQYPDYALTVTGHSLGASMAALTAA QLSATYDNVRLYTFGEPRSGNQAFASYMNDAFQVSSP ETTQYFRVTHSNDGIPNLPPADEGYAHGGVEYWSV DPYSAQNTFVCTGDEVQCCEAQGGQGVNDAHTTYFG MTSGACTW
NcFae-I	uniprot	Q9HGR3	>Q9HGR3_Feruloyl_Esterase_B MLPRTLLGLALTAATGLCASLQQVTNWGSNPTNIRMY TYVPDKLATKPAIIVALHGCGGTAPSWYSGTRLPSY ADQYGFILIYPGTPNMSNCWGVNDPASLTHGAGGDSL GIVAMVNYTIKYNADASRVYVMGTSSGGMMTNVMAA ATYPEVFEAGAAYSGVAHACFAGAASATPFSPNQTCA RGLQHTPEEWGNFVRNSYPGYTGRRPRMQIYHGLA DNLVYPRCAMEALKQWSNVLGVEFSRNVSGVPSQAY TQIVYGDGSKLVGYMGAGVGHVAPTNEQVMLKFFGLI N
FAEC_EMENI	uniprot	Q5B2G3	>Q5B2G3_Feruloyl_Esterase_C MLRAVLLPTLLAFGAFTPVHGANSPGCGKQPTLTNGV NQINGREYVLKIPDGYDPSKPHHLIFGLHWRGGNMY nvvngdsiqpwyglearaqgsaifvapnglnagwantngedvafidaimeqved dlcvdqasrfatgfswgge GMSYALACARAAEFRAVSVLSGGLISGCDGGNDPIAY LGIHGINDPVLPLDGGVTLANTFVSNNGCQPTDIGQ PASGSGGSVRTDFSGCSHPVSFIAYDGGHDGAPLGVGS SLAPDATWEFFMAA
AcFAE	NCBI	XP_001274884.1	>XP_001274884.1_Putative_Esterase MLPTILYSAILALSALTPSALAETRSSGCGKHPSLANGV IHLNGREYILKLPDRYDNNHAYHLVFGLHWR GGNMQNVANGESIQPWYGLETRAQGSTIFIAPNGKNA GWANNGGEDVAFIDAIIKQVEADLCVDQSSRFA TGFSWGGGMSYSLACSRAKQFKAVSVLSGGVISGCDG GHDPIAYLGIHGINDGVLPFNGGVGLAQKFVQN NGCQQANIGAPPSGSKSSVRTDFKGCSKPVSFIAYDGG HDSAPLGVGSSLAPDATWKFFMAA

2.2 Homology modeling of FAEs

In order to deeply investigate the structure-function relationship of the four ferulic acid esterases (AnFaeA, NcFae-I, FAEC_EMENI, and AcFAE), a three-dimensional structural model of these enzymes was constructed using homology modeling in this

study. Homology modeling is a computational method to predict the 3D structure of target proteins based on sequence similarity using homologous proteins with known structures as templates. This method can provide valuable information about protein structure and function in the absence of experimentally determined structural data.

2.2.1 SWISS-MODEL homology modeling

SWISS-MODEL is a widely used, fully automated protein structure homology modeling server that provides access through the Expasy web server [19]. The platform integrates a series of advanced modeling algorithms and optimization processes, including template search, sequence comparison, model construction, energy minimization, etc., to efficiently and accurately predict the 3D structures of proteins. The goal of SWISS-MODEL is to make protein modeling technology easily accessible and usable by all life science researchers around the world, and to promote the development of structural biology research.

For class A ferulic esterase AnFaeA, the crystal structure of ferulic esterase from Aspergillus niger (SMTL ID: 2hl6.1) was selected as a template; for class B ferulic esterase NcFae-I, carboxylic ester hydrolase from Monosporascus sp MG133 (UniProt ID: A0A0G4LQ84_9PEZI) was selected as a template; for the class C ferulic esterase FAEC_EMENI, the ferulic esterase from Aspergillus clavatus (UniProt ID: A1C9D4) was chosen as a template; for the class D ferulic esterase AcFAE (NCBI ID: XP_001274884.1), the ferulic esterase from Aspergillus clavatus (UniProt ID: A1C9D4) was used as a template.

2.2.2 Plotting Raschel diagrams to assess modeling results

In order to fully analyze the stereochemical quality of the models, the PROCHECK tool provided by the SAVES v6.0 website was used in this study[20]. PROCHECK is a widely used program for structural quality assessment of proteins, which visualizes the distribution of the main chain dihedral angles φ and ψ for each amino acid residue by drawing a Ramachandran Plot.

The Ramachandran Plot can be used to classify residues into Most favored regions, Additionally allowed regions, Generously allowed regions, and Disallowed regions, depending on the dihedral angle regions. Disallowed regions.) The higher the proportion of residues located in the Most favored regions and Additionally allowed regions, the more reasonable the main chain conformation of the protein structure and the better the stereochemical quality. In addition, PROCHECK provides statistical analysis of the overall structural geometric parameters, such as peptide bond planarity, bond length, bond angle, etc., which further assesses the structural rationality.

Using the PROCHECK tool in SAVES v6.0 website for comprehensive stereochemical quality assessment of protein structures, the reasonableness and reliability of homology modeling models can be effectively judged by analyzing the Raschel plots and overall geometrical parameters, which can provide important references for subsequent structure optimization and functional studies.

2.3 Enzyme and modeling substrate pretreatment

In order to optimize the structure of enzyme proteins and prepare them for molecular docking studies, we used a series of structural modification and optimization steps. First, the enzyme proteins were hydrogenated using the MolProbity 4.5.2 website (http://molprobity.biochem.duke.edu/). MolProbity is a widely used protein structure validation and optimization tool that can assess the quality of the structure and repair and optimize the structure, such as adding missing hydrogen atoms, flipping incorrect side-chain conformations, and so on.

Next, further preparation of the hydrogenated enzyme protein structure was carried out using AutoDock Tools software. AutoDock Tools is a comprehensive graphical interface for molecular docking that provides a range of structural preprocessing and optimization functions[21] We used the software to normalize the nonpolar hydrogen atoms of the enzyme proteins to ensure that their atom types and charge assignments conform to the requirements of molecular docking calculations. In addition, we used AutoDock Tools to add structured water molecules to the lipid bilayer region of the enzyme proteins to mimic the real state of the enzymes in a

biofilm environment. For non-standard amino acid residues present in the enzyme protein, such as modified amino acids or cofactors, we used the corresponding tools in AutoDock Tools to perform parameterization and charge assignment to ensure that they are correctly handled in molecular docking calculations. Finally, we removed the alternative back (AltB) atoms in the enzyme protein structure, which are usually derived from alternative conformations during structure determination and have no practical impact on molecular docking calculations.

For the model substrate molecules, we performed energy minimization using Chem3D software. Through energy minimization, we can obtain the low-energy conformations of the model substrate molecules, which helps to improve the accuracy and efficiency of molecular docking calculations[22].

2.4 Active site search

2.4.1 dali server structure comparison and similarity search

In order to accurately identify the active site of ferulic acid esterase, we utilized the DALI server (http://ekhidna2.biocenter.helsinki.fi/dali/) for structure comparison and similarity search. DALI is a widely used online tool that can compare the structure of a target protein with all known structures in the Protein Data Bank (PDB) and identify proteins with high structural similarity to the target protein[23]. By this method, we can identify proteins with structural similarity to ferulic acid esterase and draw on the functional and active site information of these proteins to provide important references for active site prediction of ferulic acid esterase.

The pdb files of ferulic esterase types A, B, C and D were uploaded to the DALI server respectively, and the top five proteins with structural similarity were selected for further analysis. By carefully reading the related literature of these proteins, we found that some of them have similar biological functions and sequence features with the ferulic acid esterase we studied. Based on this information, we further screened the top five proteins and finally selected for reference 1USW corresponding to type A ferulic esterase, 8IY8 corresponding to type B ferulic esterase, and 5X6S and 8IY8 corresponding to type C and D ferulic esterases. These proteins have high structural

and functional similarity to the ferulic acid esterase we studied, and their active site information is an important reference value for predicting the active site of ferulic acid esterase.

2.4.2 Multiple sequence comparison and phylogenetic tree construction

When performing evolutionary analysis of protein sequences, the presence of signal peptides may have an impact on the results. Studies have shown that sequence comparison after removing signal peptides can provide more accurate and reliable conclusions for evolutionary analysis[24]. In order to obtain a higher quality homology modeling model, SignalP 6.0 server (https://services.healthtech.dtu.dk/services/SignalP-6.0/) was used in this study to predict signal peptides for each sequence[25]. SignalP 6.0 is a deep learning-based signal peptide prediction tool, which significantly improves the prediction accuracy and speed compared to previous versions. With this server, the signal peptide regions of each sequence were successfully identified and they were removed from the downloaded sequence.

In order to further analyze the similarity and conservation of these proteins with ferulic acid esterase at the sequence level, the MUSCLE tool in MEGA11 software was used. MEGA software is a widely used molecular evolutionary analysis software, which integrates a variety of sequence comparison and phylogenetic analysis algorithms for the study of evolutionary biology[26].MUSCLE is a fast and accurate MUSCLE is a fast and accurate multiple sequence comparison algorithm, which can identify conserved regions and key residues in the sequences.

With the purpose of exploring the evolutionary relationships and functional diversity of the ferulic acid esterase family in greater depth, a comprehensive phylogenetic tree was constructed using relevant sequences retrieved from the DALI server. The Neighbor-Joining Tree algorithm in MEGA11 software was used. Neighbor-Joining (NJ) is a commonly used method to construct a phylogenetic tree, which is based on a distance matrix, and ultimately constructs an unrooted tree by continuously selecting the two nodes with the closest distances for clustering[27]. Phylogenetic tree is a widely used bioinformatics tool, which can visualize the

evolutionary relationships and affinities between different biological sequences. By analyzing the clustering and branching patterns of ferulic acid esterase family members in the phylogenetic tree, we can infer their evolutionary history and functional differentiation process.

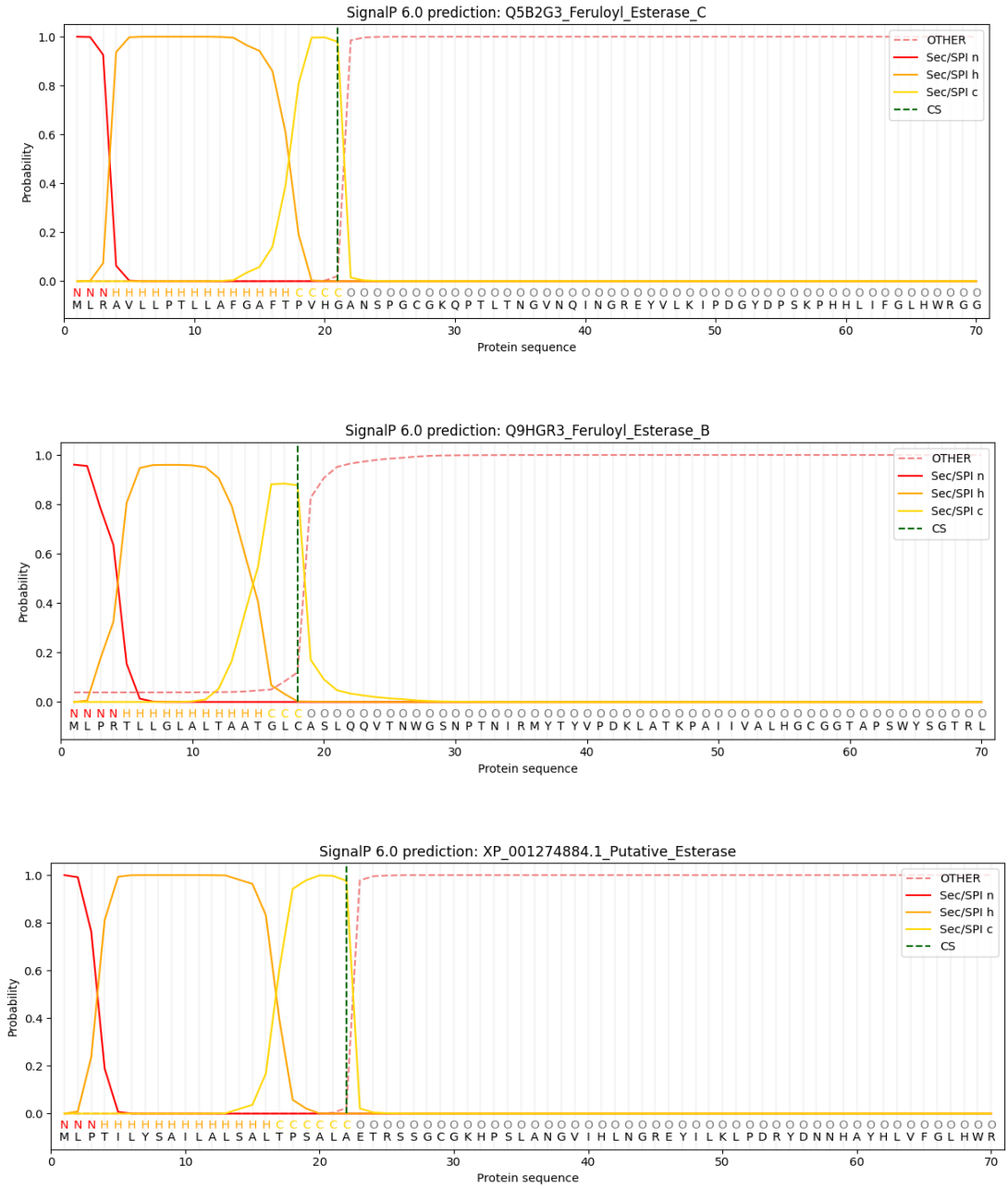
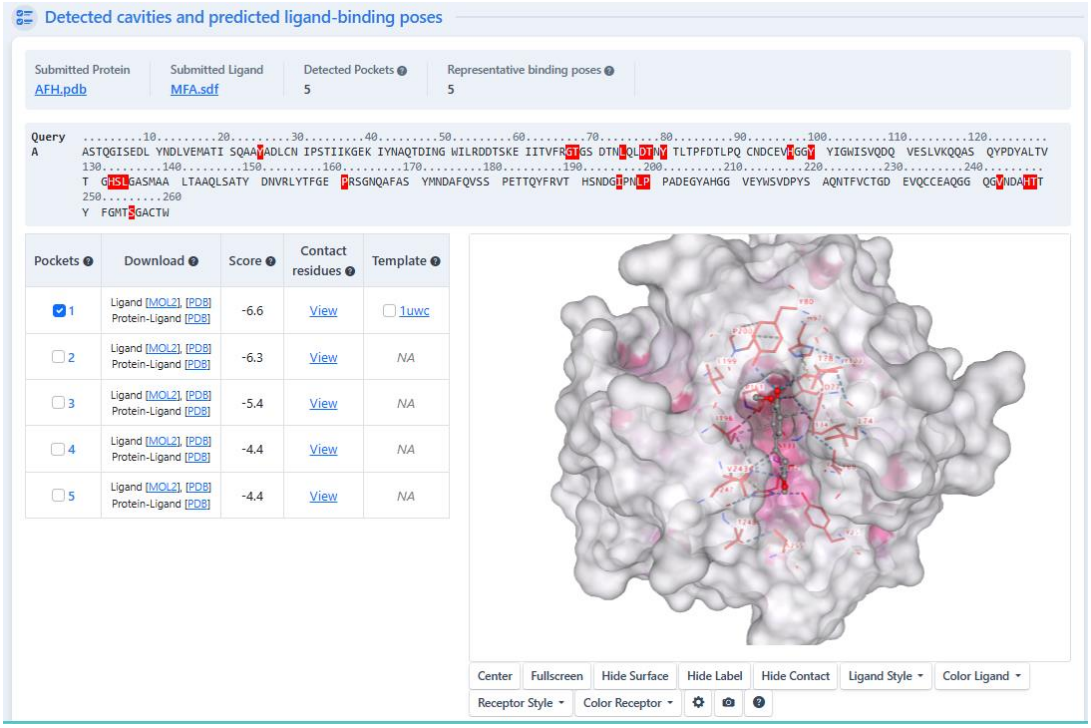


Figure 2.1 – Signal peptide finding

2.4.3 Prediction of CB-Dock2 ligand binding sites

To further validate and refine our predicted ferulic acid esterase active site, we introduced the CB-Dock2 website (https://cadd.labshare.cn/cb-dock2/php/blinddock.php) for protein-ligand blind docking analysis. CB-Dock2 is a computationally based blind docking tool that improves and extends the previous version of CB-Dock by integrating a variety of advanced algorithms and databases to more accurately and comprehensively predict potential ligand binding sites on proteins[28].

The built-in perl script of CB-Dock2 will automatically process these files and call the latest version (1.2.0) of AutoDock Vina to perform template-independent blind docking calculations. At the same time, the tool will also utilize the homologous template information in the BioLip database to perform template-based blind docking analysis. By combining the results of these two methods, CB-Dock2 can give a prediction of the most likely ligand binding site on ferulic acid esterase.



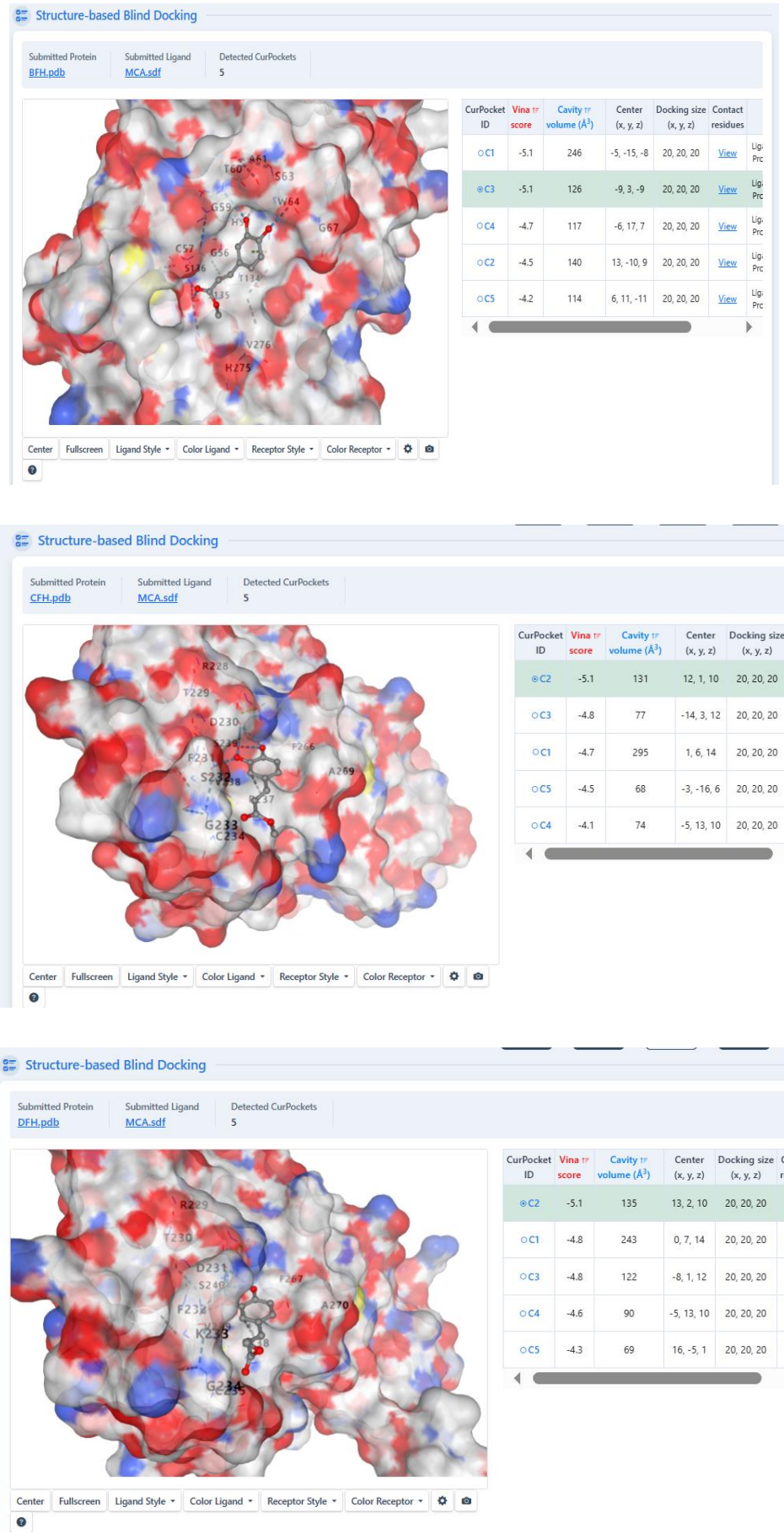


Figure 2.2 – CB-Dock2 docking results

2.4.4 Active site determination

Literature has shown that by studying the structural biology of ferulic esterases from different sources, it has been found that their active site is usually composed of the following key components: the catalytic triad, the oxygen anion pore, the hydrophobic pocket, and the substrate-binding pocket.

The catalytic triad consists of serine (Ser), histidine (His), and aspartic acid (Asp), and the serine is located in a universally conserved pentapeptide sequence, "GXSXG", which forms part of the active site of the enzyme.

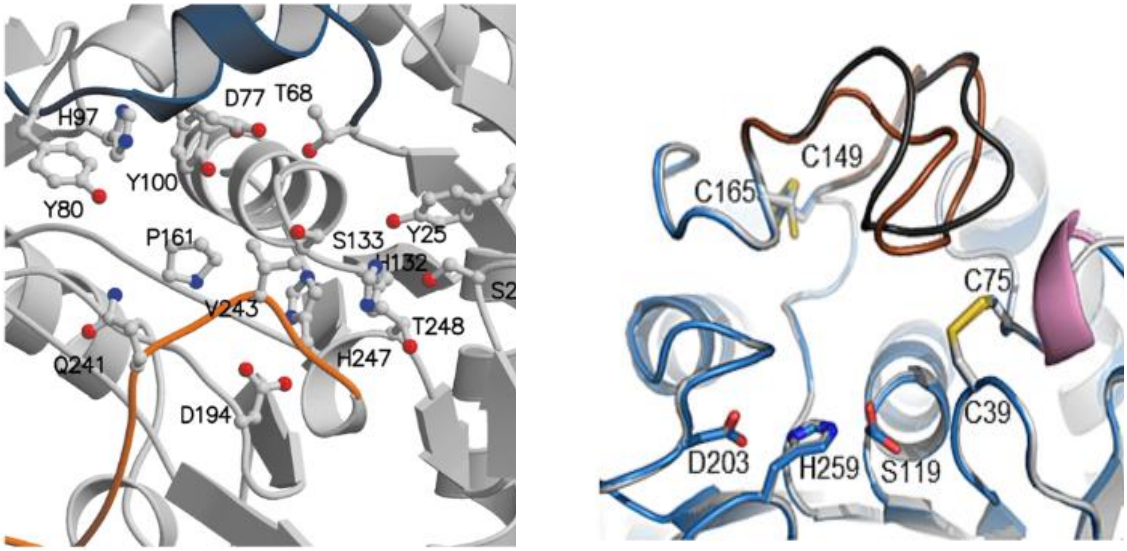


Figure 2.3 – Literature sites related to A and B enzymes[29-30]

Through the pymol synthesis results, we identified and carried out the observation of the catalytic triplex positions of the four classes of ferulic acid esterase ABCD, the corresponding S-D-H catalytic triplexes of the four classes of enzymes are shown in the following table, and the observation of the triplex positions is shown in the figure.

Table 2.2 – **Catalytic triad positions**

Type of enzyme	Name of enzyme	Catalytic triad position
A	AnFaeA	S133,D194,H132
B	NcFae-I	S135,D219,H275
C	FAEC_EMENI	S142,D190,H186
D	AcFAE	S144,D192,H188

"

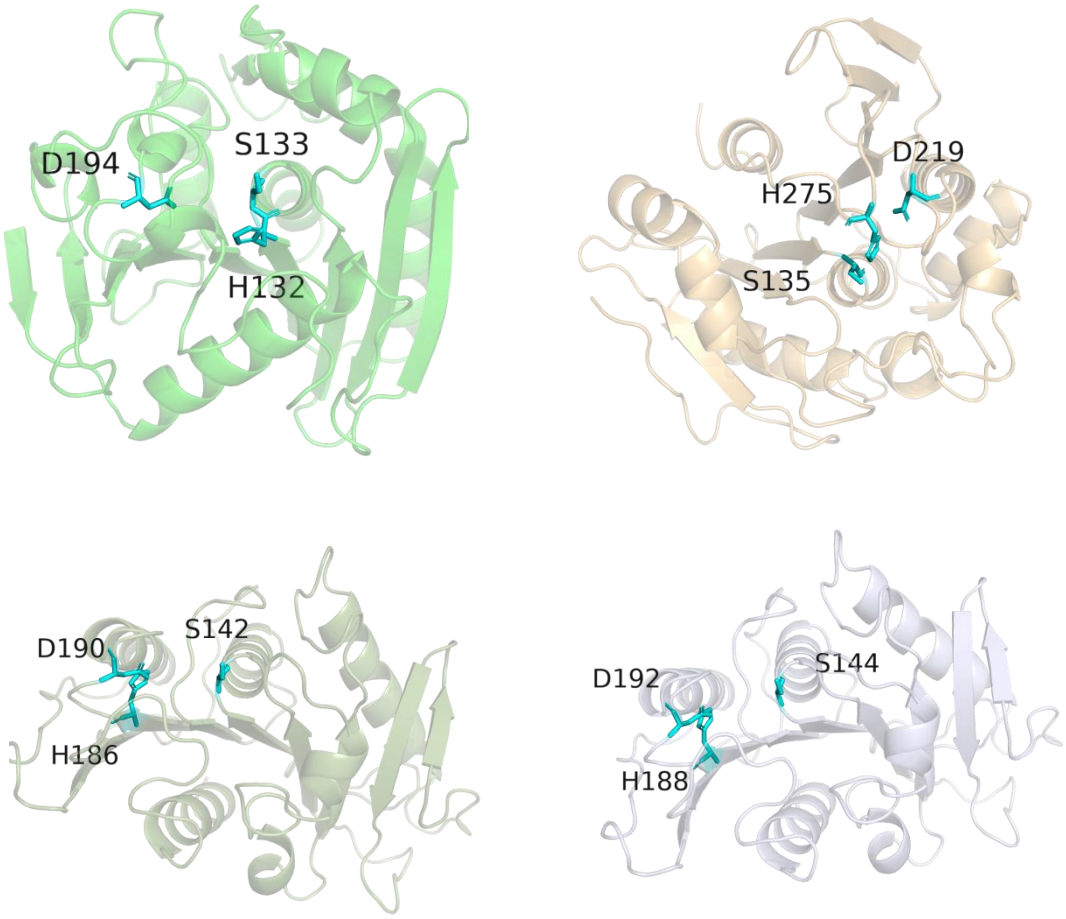


Figure 2.4 – Catalytic triad positions

In order to more accurately represent the active pocket regions of the C and D enzymes, which temporarily lack similar structural literature, residues within the 1.5 Å region around the catalytic triad and spatially located close to the catalytic triad were selected as active pocket regions.

Finally the active pocket residues identified in this study to carry out further research are shown in the following table.

Table 2.3 – **Active pocket residues**

Type of enzyme	Name of enzyme	Active pocket residues
A	AnFaeA	Y25T68,D77,Y80,H97,Y100,H132,S133,P161,D194,Q241, V243,H247,T248,S255
B	NcFae-I	C57,C92,S135,S136,C182,H275,D219
C	FAEC_EMENI	F141,S142,I185,H186,G187,N189,D190,P191
D	AcFAE	F143,S144,I187,H188,G189,N191,D192,G193

2.5 Model substrate docking based on enzyme active sites

Systematic molecular docking studies were carried out using AutoDock software. It is known that class A ferulic esterase AnFaeA has binding activity only with the model substrates MFA, MpCA, and MSA, class B ferulic esterase NcFae-I has binding activity only with the model substrates MCA, MFA, and MpCA, and class C ferulic esterase FAEC_EMENI was not available, and thus its docking with all four model substrates was performed, and class D ferulic esterase AcFAE had binding activity only with the model substrates MCA, MFA, and MpCA. In this study, only docking experiments with active model substrates were performed, and the parameters of docking are shown in the following table.

Table 2.4 – **Docking parameters**

Class of enzyme	Name of enzyme	npts	Spacing	Gridcenter
Gridcenter	AnFaeA	74 54 72	0.375	41.4 2.8 19.8
B	NcFae-I	53 41 59	0.375	6.95 8.55 -8.6
C	FAEC_EMENI	74 36 56	0.375	5.35 3.45 -9.65
D	AcFAE	56 35 54	0.375	8.85 3.3 -9.05

The number of docking was 20 and the result with the lowest binding energy was retained.

2.6 Mapping based on docking results

2.6.1 2D mapping analysis

LigPlus+ is a powerful molecular docking visualization software, which can generate high-quality 2D interaction diagrams and clearly show the details of ligand-receptor interactions, including hydrogen bonding, hydrophobic interactions, π-π stacking and so on.

2.6.2 3D mapping analysis

On the basis of the 2D analysis, we further used PLIP website + PyMOL software to visualize the molecular docking results in 3D. PLP website can easily and quickly identify non-covalent interactions between biomolecules and their ligands .

With PLIP, we can look for the interaction force between the enzyme and the model substrate in the docking results. With PyMOL, we can visualize the spatial position and orientation of the ligand in the receptor binding pocket and analyze the interaction network formed by the ligand with the surrounding amino acid residues.

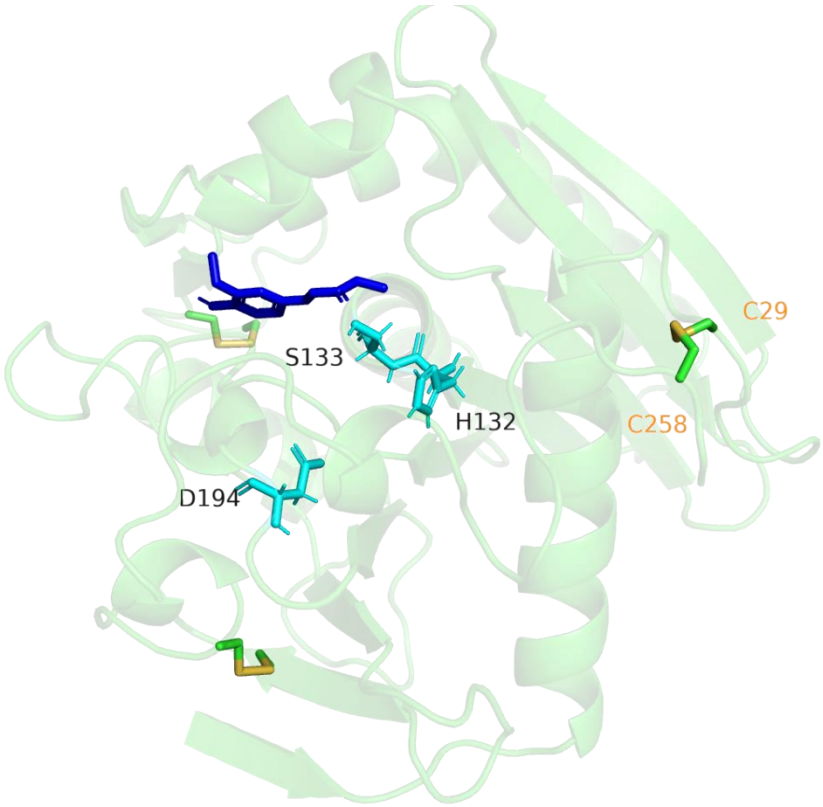
2.7 Study of the effect of disulfide bonds on the binding activity around the enzyme active site

For the two groups of AB enzyme, since the disulfide bond is close to the active site, this may mean that the disulfide bond has a direct effect on the structure and function of the active site. Therefore, it is more meaningful to perform the experiment of breaking disulfide bond for these two groups, because then the change of enzyme activity after breaking the disulfide bond can be observed, which can infer the effect of disulfide bond on enzyme activity.

For the CD enzyme two groups, since the disulfide bond is far away from the active site, this could mean that the disulfide bond has less of an effect on the active site. In this case, experiments with added disulfide bonds were performed to verify if the alteration of the disulfide bond would cause enough changes to the overall structure of the enzyme to affect the active site and enzyme activity.

2.7.1 Knockdown of disulfide bonds in group AB

Disulfide bonds in the vicinity around the binding site of the model substrate were observed by pymol. For group A ferulic esterases, the disulfide bonds formed by C29-C258 and C227-C234 were chosen to be knocked down by grouping mutation of C258 and C234 to alanine Ala, respectively, and for group B ferulic esterases, the disulfide bonds formed by C57-C92 were chosen to be knocked down by mutating C57 to alanine Ala. After homology modeling using SWISS-MODEL, the same template was used for class A and B as in the previous modeling, and the scores were GMQE 0.98 for both groups of class A enzymes and GMQE 0.94 for class B enzymes, which proved that the results of homology modeling were good), and then hydrotreating by MolProbity, and molecular docking was performed by using autodock 4.2.6. for molecular docking.



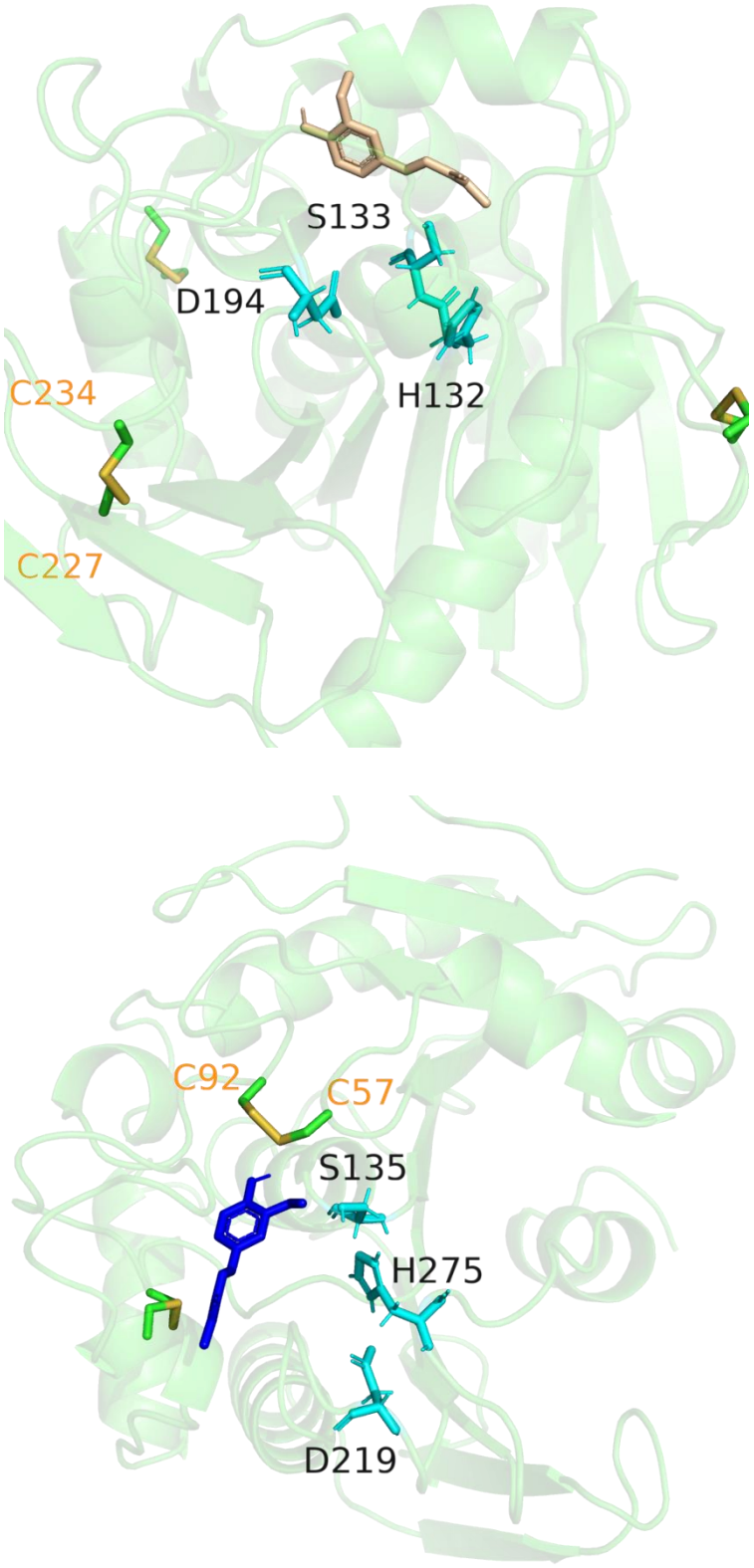


Figure 2.5 – Disulfide bond 3D position

2.7.2 Mutating CD group residues to form disulfide bonds

Disulfide bonds in the vicinity around the binding site of the model substrate were searched for by Disulfide by Design2[32], and W109 and M149 of mutant class C and

D ferulic acid esterases were chosen to form the disulfide bonds, which were hydrogenated and pretreated by MolProbity, and molecular docking was carried out using autodock 4.2.6.

2.8 Kinetic simulation

2.8.1 Construction of the initial simulation system in aqueous solution

In this study, we used GROMACS2022.2. GROMACS is a widely used MD simulation tool that has been applied and validated in the study of numerous biomolecular systems . The first step of the simulation is to prepare the initial conformation of the complex, which is a critical step that directly affects the accuracy and reliability of subsequent simulations.

In order to obtain a high-quality initial conformation, we first treated the protein and the small molecule substrate separately. Proteins were simulated using the CHARMM36 force field, a widely used molecular dynamics simulation force field specifically designed to simulate a wide variety of biomolecules, including proteins, nucleic acids, and lipids . The CHARMM36 force field has been validated in numerous studies to accurately characterize the structure and dynamics of proteins.

For small molecule substrates, we used Avogadro software for hydrogenation and converted them into mol2 format files. However, the CHARMM36 force field may not have predefined parameters for small molecules, and to address this issue, we utilized the CHARMM General Force Field (CGenFF) web program (https://app.cgenff.com/) to generate force field parameters for small molecules. CGenFF is part of the CHARMM force field family and is specifically designed to handle the parameterization of small molecules. It uses a chemical topology-based approach to predict the force field parameters of various small molecule structures, providing a reliable basis for simulations[31].

In this study, the simulation box is set as a cube and the minimum distance between the molecules and the box boundary is set to 1.0 nm (-d 1.0). This setting helps to ensure that the molecules in the simulation have enough space to move and

avoid immediate collisions with the box boundary, thus improving the reliability of the simulation results.

In constructing the water environment, the SPC216 water model, a simple water model widely used in molecular dynamics simulations, was used in this study. The SPC216 water model is able to reproduce the structural and dynamical properties of water better, while it is computationally efficient and suitable for large-scale molecular dynamics simulations. In order to maintain the electroneutrality of the system or adjust the ion concentration, an appropriate amount of ions was added to the system in this study.

After the initial construction of the simulation system was completed, the first energy minimization process was carried out to reduce the possible high energy defects in the initial structure. Energy minimization is an important step in molecular dynamics simulation and helps to obtain a more reasonable and stable initial structure. In the first energy minimization, more conservative parameter settings were used to ensure the stability of the system. The energy minimization process uses steepest descent minimization as the integration algorithm, setting the stopping of iteration when the maximum force is less than 10.0 kJ/mol, an energy step size of 0.01, and a maximum number of iteration steps of 50,000. These parameter settings can effectively optimize the system structure and eliminate high-energy defects within a reasonable computation time.

2.8.2 Molecular dynamics simulation

The second energy minimization is usually performed after adding ions to the system or making other structural modifications to ensure that the new structure is still at an energy minimum. Compared to the first minimization, the second minimization employs some more optimized parameter settings to further improve the accuracy and stability of the simulation.

In the second energy minimization, the truncation radius of the short-range van der Waals (vdW) and electrostatic forces was increased from 1.0 to 1.2 nm. This adjustment helps to consider the intermolecular interactions more comprehensively

and improve the accuracy of the simulation results. Meanwhile, in order to deal with long-range electrostatic interactions more efficiently, the PME (Particle Mesh Ewald) method was used in this study. By decomposing the electrostatic interactions into two parts, short-range and long-range, and using the Fourier transform to deal with the long-range part, the PME method is able to significantly reduce the computational cost while ensuring the computational accuracy.

The two energy minimization settings are designed to adapt to the different needs in the simulation process and improve the stability and accuracy of the simulation. The first minimization mainly ensures the stability of the initial structure and eliminates possible high energy defects. The second minimization, on the other hand, optimizes for the modified structure and further improves the starting quality of the simulation. These two steps are crucial for subsequent molecular dynamics production runs, as a good starting structure reduces instability during simulations and improves the reliability of results.

In order to gain a deeper understanding of the mechanism of protein-ligand interactions, we performed a series of molecular dynamics simulations of the target complexes.

First, we equilibrated the system under the NVT system, in which the volume (V) and temperature (T) were kept constant, the number of simulation steps was 50,000, and the saving frequencies of energy and coordinates were set to every 1.0 ps, respectively, for subsequent analysis.

After the NVT equilibrium, we switched the system to the NPT system. The reference pressure was set to 1.0 bar, which matched the isothermal compressibility of water at room temperature. The NPT equilibrium stage took over the atomic velocity of the NVT equilibrium stage to ensure the continuity of the system between different systems.

Finally, we performed a long time molecular dynamics simulation to obtain the conformational and kinetic information of the protein-ligand complexes in the equilibrium state. The total number of simulation steps was increased to 5,000,000, and the corresponding simulation time was 10,000 ps (10 ns).

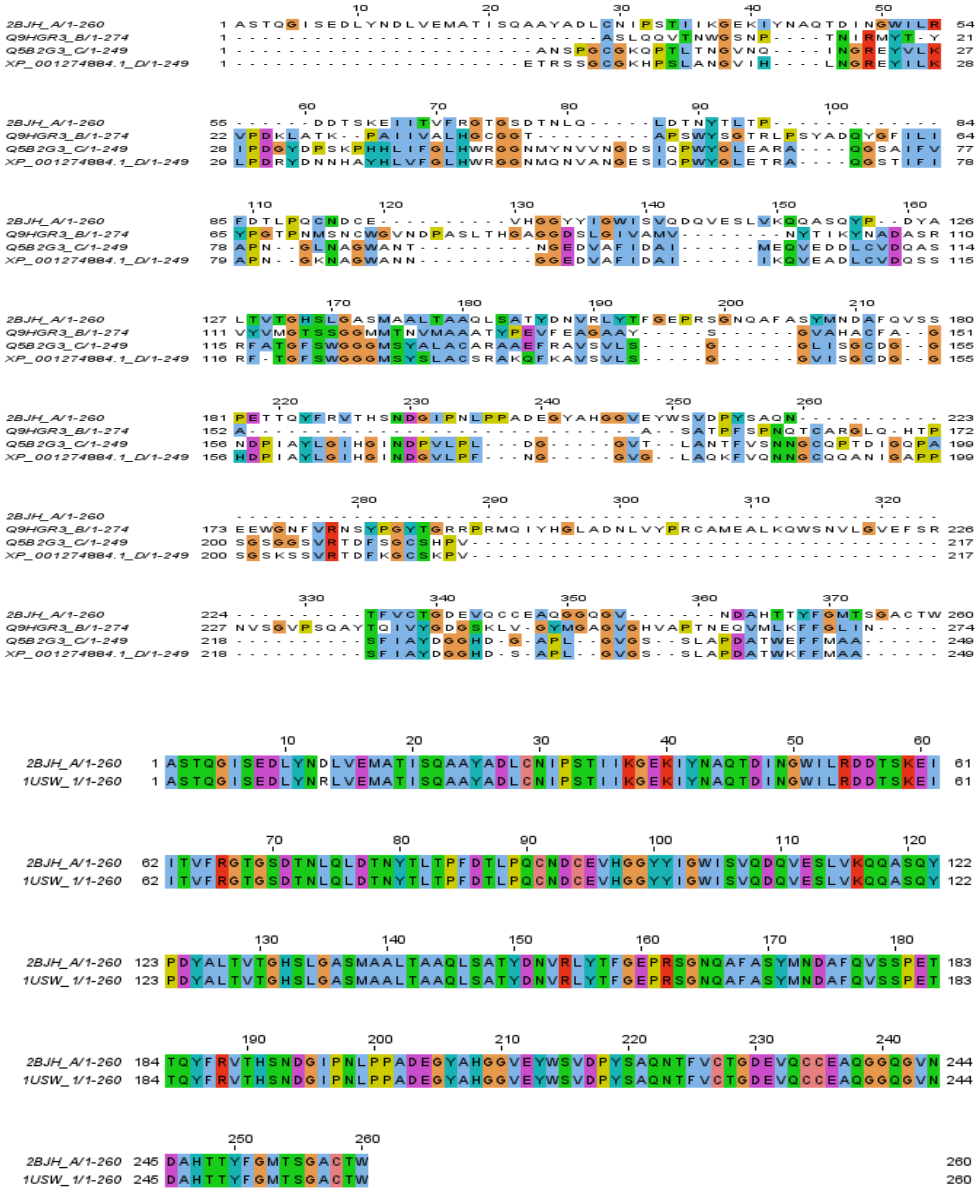
CHAPTER 3

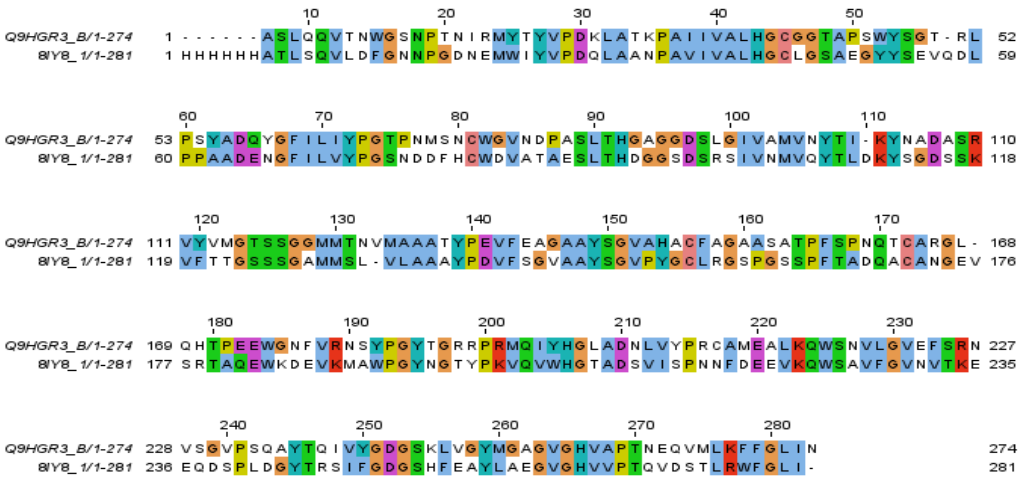
EXPERIMENTAL PART

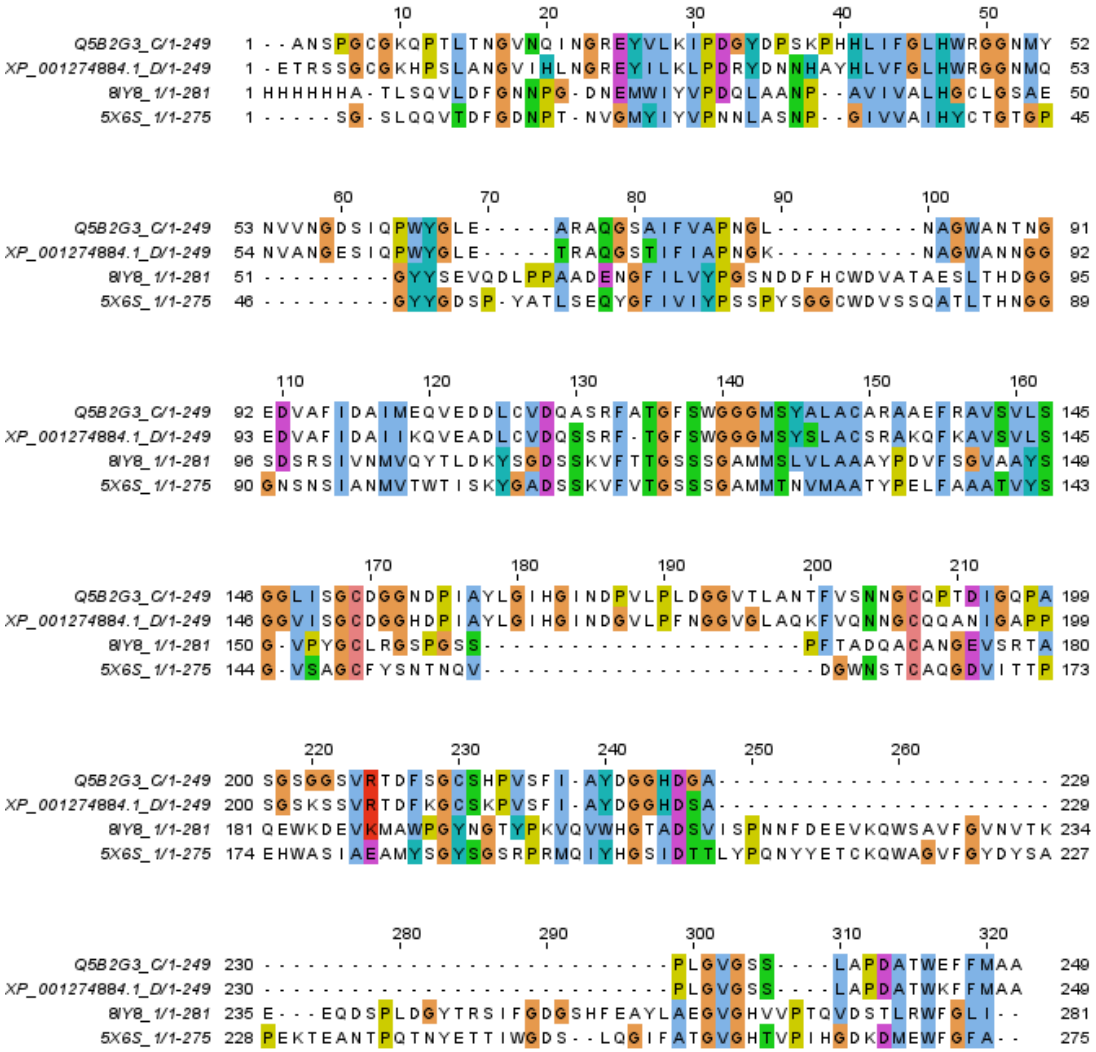
3.1 Comparative sequence analysis

3.1.1 Sequence analysis

The results of multiple sequence alignment using CLUSTAL Omega 1.2.2 (http://www.clustal.org/omega/) and sequence visualization with the Jalview software are as follows.




Q9HGR3_B/1-274 1 ------ASLQQVTNWGSNPTNIRMYTYVPDKLATKPAIIVALHGCGGTAPSWYSGT-RL 52
8IY8_1/1-281 1 HHHHHHATLSQVLDFGNNPGDNEMWIYVPDQLAANPAVIVALHGCLGSAEGYYSEVQDL 59
Q9HGR3_B/1-274 53 PSYADQYGFILIYPGTPNMSNCWGVNDPASLTHGAGGDSLGIVAMVNYTI-KYNADASR 110
8IY8_1/1-281 60 PPAADENGFILVYPGSNDDFHCWDVATAESLTHDGGSDSRSIVNMVQYTLDKYSGDSSK 118
Q9HGR3_B/1-274 111 VYVMGTSSGGMMTNVMAAATYPEVFEAGAAYSGVAHACFAGAASATPFSPNQTCARGL- 168
8IY8_1/1-281 119 VFTTGSSSGAMMSL-VLAAAYPDVFSGVAAYSGVPYGCLRGSPGSSPFTADQACANGEV 176
Q9HGR3_B/1-274 169 QHTPEEWGNFVRNSYPGYTGRRPRMQIYHGLADNLVYPRCAMEALKQWSNVLGVEFSRN 227
8IY8_1/1-281 177 SRTAQEWKDEVKMAWPGYNGTYPKVQVWHGTADSVISPNNFDEEVKQWSAVFGVNVTKE 235
Q9HGR3_B/1-274 228 VSGVPSQAYTQIVYGDGSKLVGYMGAGVGHVAPTNEQVMLKFFGLIN 274
8IY8_1/1-281 236 EQDSPLDGYTRSIFGDGSHFEAYLAEGVGHVVPTQVDSTLRWFGLI- 281


Q5B2G3_C/1-249 1 --ANSPGCGKQPTLTNGVNQINGREYVLKIPDGYDPSKPHHLIFGLHWRGGNMY 52
XP_001274884.1_D/1-249 1 -ETRSSGCGKHPSLANGVIHLNGREYILKLPDRYDNNHAYHLVFGLHWRGGNMQ 53
8IY8_1/1-281 1 HHHHHHA-TLSQVLDFGNNPG-DNEMWIYVPDQLAANP--AVIVALHGCLGSAE 50
5X6S_1/1-275 1 -----SG-SLQQVTDFGDNPT-NVGMYIYVPNNLASNP--GIVVAIHYCTGTGP 45
Q5B2G3_C/1-249 53 NVVNGDSIQPWYGLE-----ARAQGSAIFVAPNGL----------NAGWANTNG 91
XP_001274884.1_D/1-249 54 NVANGESIQPWYGLE-----TRAQGSTIFIAPNGK----------NAGWANNGG 92
8IY8_1/1-281 51 ---------GYYSEVQDLPPAADENGFILVYPGSNDDFHCWDVATAESLTHDGG 95
5X6S_1/1-275 46 ---------GYYGDSP-YATLSEQYGFIVIYPSSPYSGGCWDVSSQATLTHNGG 89
Q5B2G3_C/1-249 92 EDVAFIDAIMEQVEDDLCVDQASRFATGFSWGGGMSYALACARAAEFRAVSVLS 145
XP_001274884.1_D/1-249 93 EDVAFIDAIIKQVEADLCVDQSSRF-TGFSWGGGMSYSLACSRAKQFKAVSVLS 145
8IY8_1/1-281 96 SDSRSIVNMVQYTLDKYSGDSSKVFTTGSSSGAMMSLVLAAAYPDVFSGVAAYS 149
5X6S_1/1-275 90 GNSNSIANMVTWTISKYGADSSKVFVTGSSSGAMMTNVMAATYPELFAAATVYS 143
Q5B2G3_C/1-249 146 GGLISGCDGGNDPIAYLGIHGINDPVLPLDGGVTLANTFVSNNGCQPTDIGQPA 199
XP_001274884.1_D/1-249 146 GGVISGCDGGHDPIAYLGIHGINDGVLPFNGGVGLAQKFVQNNGCQQANIGAPP 199
8IY8_1/1-281 150 G-VPYGCLRGSPGSS----------PFTADQACANGEVSRTA 180
5X6S_1/1-275 144 G-VSAGCFYSNTNQV----------DGWNSTCAQGDVITTP 173
Q5B2G3_C/1-249 200 SGSGGSVRTDFSGCSHPVSFI-AYDGGHDGA---------- 229
XP_001274884.1_D/1-249 200 SGSKSSVRTDFKGCSKPVSFI-AYDGGHDSA---------- 229
8IY8_1/1-281 181 QEWKDEVKMAWPGYNGTYPKVQVWHGTADSVISPNNFDEEVKQWSAVFGVNVTK 234
5X6S_1/1-275 174 EHWASIAEAMYSGYSGSRPRMQIYHGSIDTTLYPQNYYETCKQWAGVFGYDYSA 227
Q5B2G3_C/1-249 230 ----------PLGVGSS----LAPDATWEFFMAA 249
XP_001274884.1_D/1-249 230 ----------PLGVGSS----LAPDATWKFFMAA 249
8IY8_1/1-281 235 E---EQDSPLDGYTRSIFGDGSHFEAYLAEGVGHVVPTQVDSTLRWFGLI-- 281
5X6S_1/1-275 228 PEKTEANTPQTNYETTIWGDS--LQGIFATGVGHTVPIHGDKDMEWFGFA-- 275

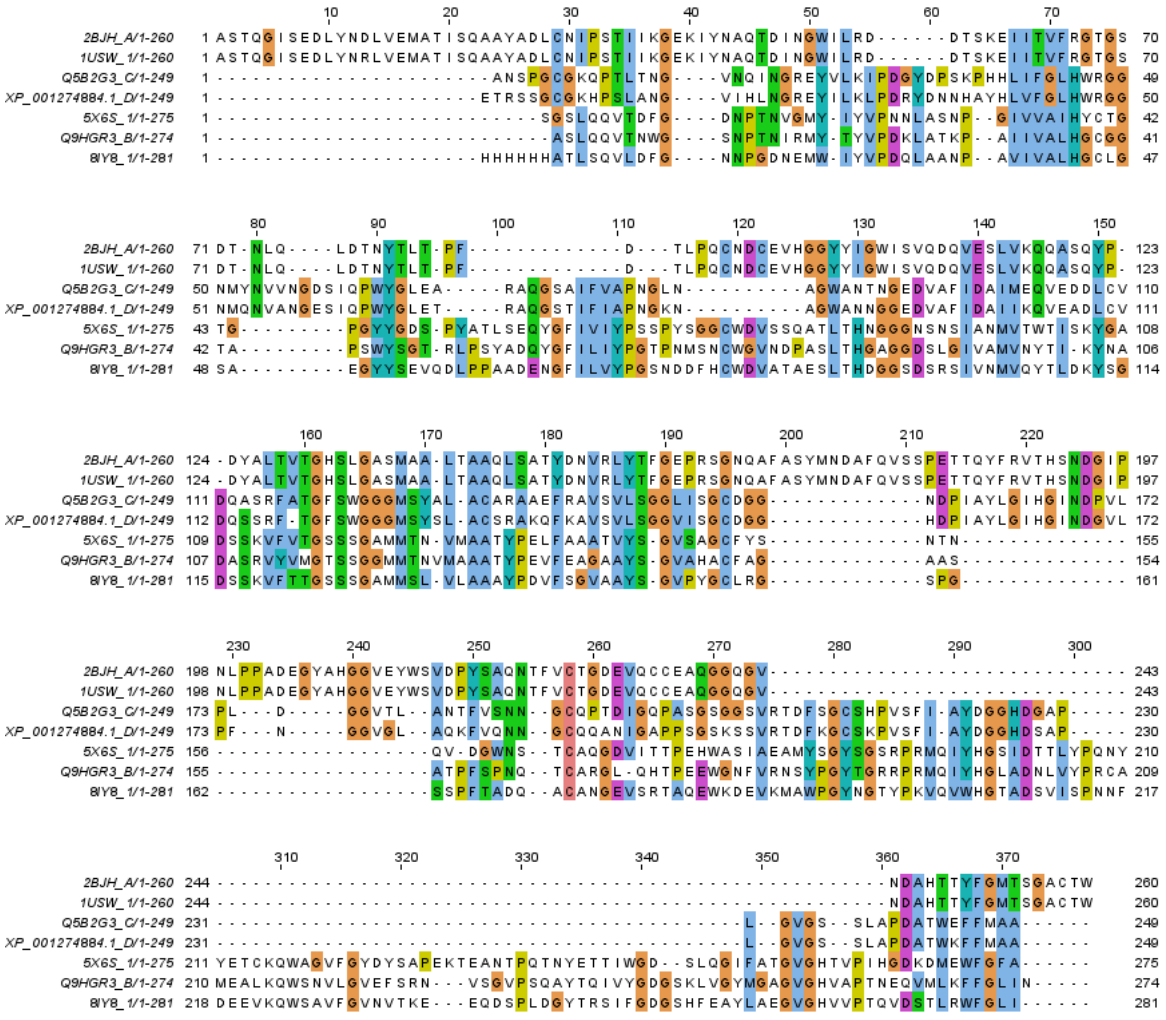


Figure 3.1 – Sequence comparison results

Class A ferulic acid esterase and its similar proteins showed high sequence conservation, especially in the amino acid residues near the active site of the enzyme. This suggests that class A enzymes may perform similar biochemical functions in different species, and their key catalytic mechanisms may have been effectively conserved during the evolutionary process. This high degree of sequence similarity suggests that class A ferulic acid esterases may assume crucial physiological roles in organisms.

Second, class B ferulic acid esterases exhibit some degree of sequence variability in certain key regions. Such differences may be closely related to the substrate

recognition and catalytic properties that characterize class B enzymes. We hypothesize that these sequence variants may be the adaptive results of class B enzymes to specific substrates in different organisms during the evolutionary process. Through the comparative analysis, we were able to identify the key amino acid sites that may determine the unique substrate specificity of class B enzymes, which provides important clues for the subsequent in-depth study.

The CD feruloyl esterases showed significant sequence differences from other classes of enzymes, especially in the regions that might be closely related to the enzyme functions. This finding attracts our great attention because it suggests that CD-like enzymes may perform some special biochemical functions in organisms. Through comparative analyses, we have identified some amino acid fragments that may be crucial for the unique functions of CD-like enzymes, and pointed out the direction for subsequent structural biology and biochemical studies.

Through comprehensive comparative analysis, we found that the Ser-Asp-His catalytic triad is strongly conserved among all classes of ferulic acid esterases. This result confirms the universal importance of this catalytic triad in the catalytic mechanism of ferulic acid esterases, and provides a key clue to elucidate the molecular mechanism of this enzyme family. Meanwhile, this finding also provides an important theoretical guidance for the modification and optimization of ferulic acid esterase by means of directed evolution.

3.1.2 Phylogenetic tree analysis

In this study, we used the Neighbor-Joining Tree algorithm in the MEGA11 software to perform phylogenetic analysis of the obtained ferulic acid esterase sequences in order to explore the evolutionary relationship between them. By constructing the phylogenetic tree, we can clearly see the kinship relationship between different classes of ferulic acid esterases.

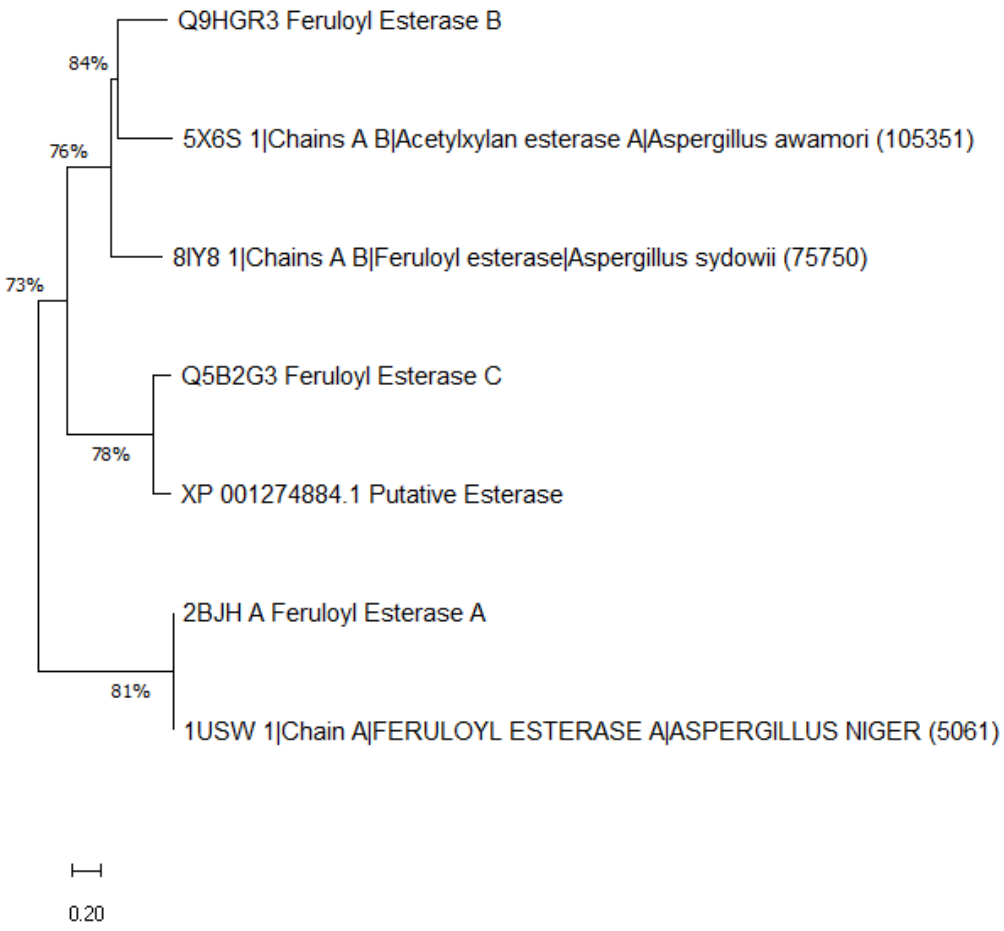


Figure 3.2 – Phylogenetic tree

It is noteworthy that the class B feruloyl esterase obtained in this study was clustered with Acetylxylan esterase from Aspergillus awamori and Feruloyl esterase from Aspergillus sydowii to form an independent evolutionary branch. This branch has high statistical support. The high statistical support rates of 84% and 76% for this branch indicate that they are evolutionarily close to each other. The above studies suggested that although the strains originated from different species, there were some structural and functional similarities between them.

The results show that the obtained C-type ferulic acid esterase and D-type ferulic acid esterase XP001274884.1 also constitute a kind of clustering with the following statistical significance:

It has seventy-eight percent support. It is hypothesized that type C and D feruloyl esterases have the same evolutionary origin, and their differentiation process most likely took place within a relatively short period of time.

In contrast, type A feruloyl esterase clustered with Feruloyl esterase from Aspergillus niger to form another independent evolutionary branch, which was supported by 81%. This result suggests that class A feruloyl esterase may be more similar in sequence and structure to the enzyme of Aspergillus niger, and it is likely that they are more closely related through a similar evolutionary process.

It was shown that the various classes of ferulic acid esters are evolutionarily significantly differentiated. Class C and D ferulic acid esterases may share a common evolutionary origin, while class A ferulic acid esterases are more closely related to their counterparts in other species of fungi. These results provide important clues for understanding the molecular evolution of ferulic acid esterases, and also lay the foundation for subsequent comparative genomics and protein engineering studies. The combination of A-type feruloyl esterase with feruloyl esterase from Aspergillus niger constitutes a new type of feruloyl esterase.

This is a new evolutionary direction, which was supported by 81%. This result suggests that the enzyme systems of type A ferulic esterase and Aspergillus niger are sequentially identical.

Tectonically, it is likely that they are more closely related through a similar evolutionary process.

It was shown that the various classes of ferulic acid esters are evolutionarily significantly differentiated. Type C, D ferulic acid esterases

Type A ferulic acid esterases share the same evolutionary origin compared to the other types of strains. Those

This study will lay the foundation for revealing the molecular evolutionary mechanism of ferulic acid esterase and for further comparative genomic and proteomic studies.

3.2 Homology modeling analysis

3.2.1 Homology modeling structure analysis

The structures obtained from homology modeling of AnFaeA, NcFae-I, FAEC_EMENI, XP_001274884.1 are shown in Fig. By comparing and analyzing the homology modeling results of the ferulic acid esterase ABCD, it is found that two enzymes, A and B, have obvious structural similarity through homology modeling of the ABCD enzyme system, and it is speculated that the two enzymes share a similar catalytic mechanism and substrate specificity. catalytic mechanism and substrate specificity. Similarly, the structural models of C and D enzymes also show a certain degree of similarity, even though their overall shapes are different from those of A and B. The structural models of A and B enzymes are also similar to those of C and D enzymes.

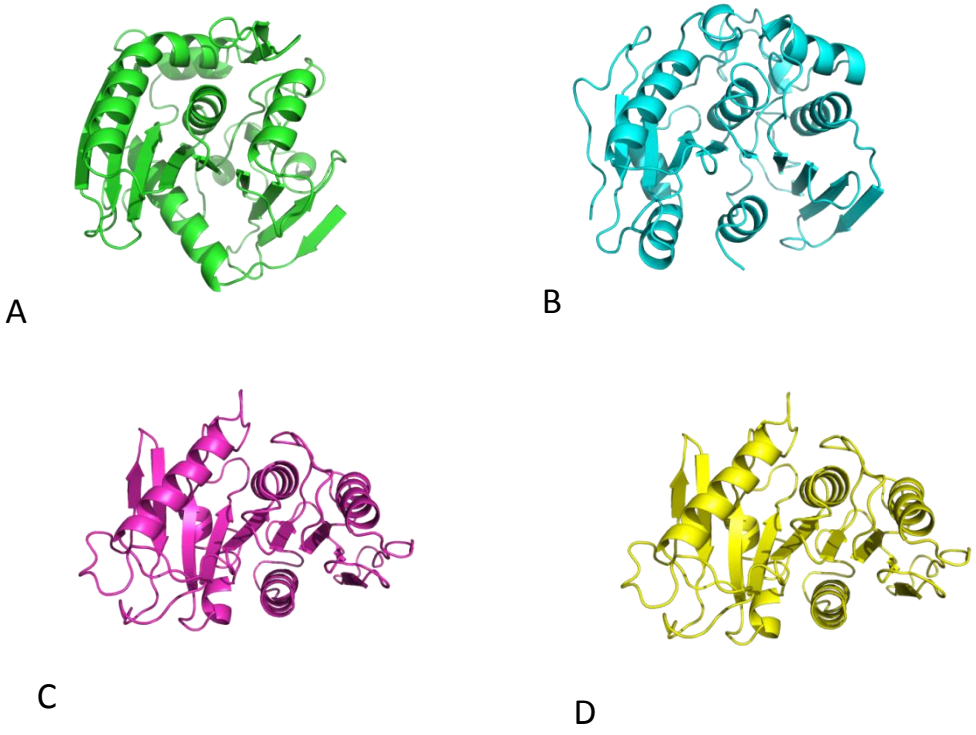


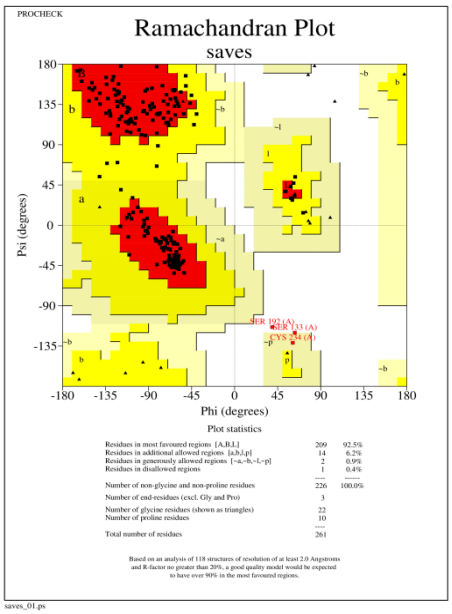
Figure 3.3 – Homology modeling results

The three-dimensional structures of each enzyme show unique folding patterns, suggesting that they may have different active sites and substrate specificities.

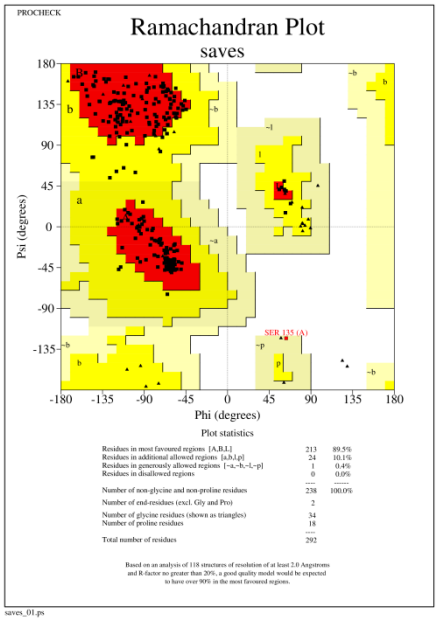
Second, although they belong to the same family, the spatial conformational differences among the four enzymes suggest that they may have adapted to different biological functions during evolution. For example, enzyme A has a relatively compact structure, which may imply that it has a high substrate affinity or is designed to function in smaller organelles. In contrast, enzyme B has a more open structure, which may help them interact with larger substrates or multimers.

3.2.2 Raschel diagram analysis

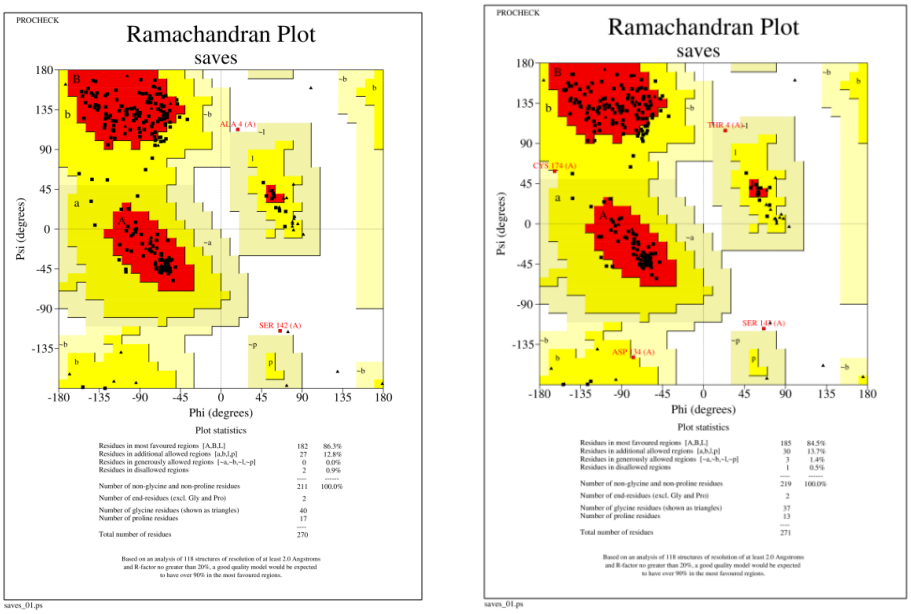
The dihedral angle distribution of most residues is within the most preferred red region (82.9%, 83.5%, 86.9%, and 84.1%, respectively), as well as the allowed yellow region, which suggests that the overall conformation of the constructed model is consistent with the known protein structure, and that overall, the constructed model displays a good distribution of dihedral angles on the whole.



A



B



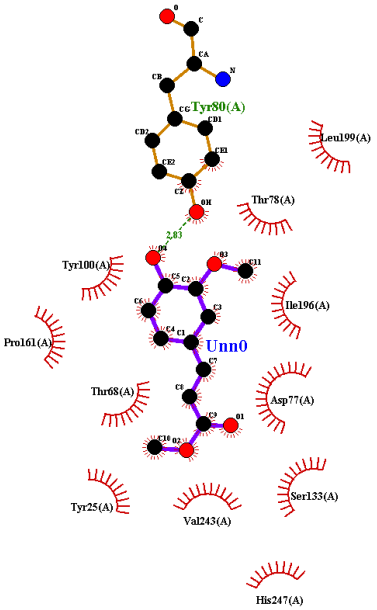
C D

Figure 3.4 – Homology modeling Raschach diagram analysis

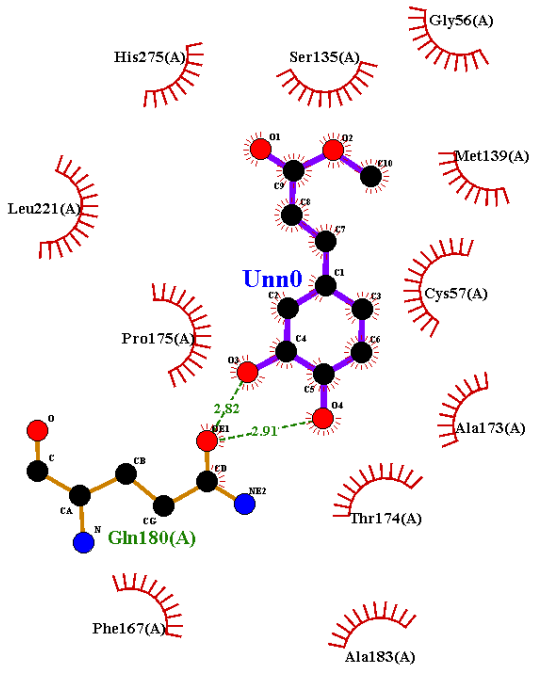
3.3 Analysis of molecular docking results

3.3.1 Docking results graphing

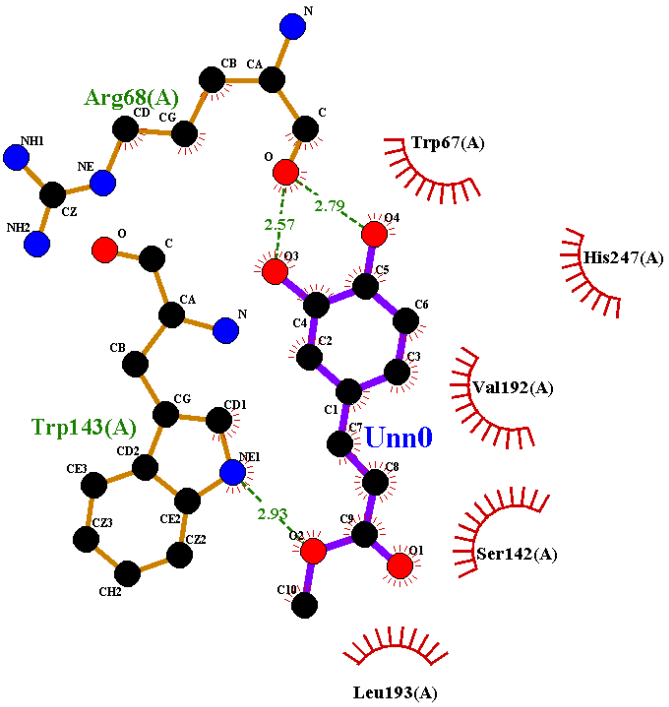
The docking results reveal the detailed interaction pattern of the substrate with each enzyme active site.



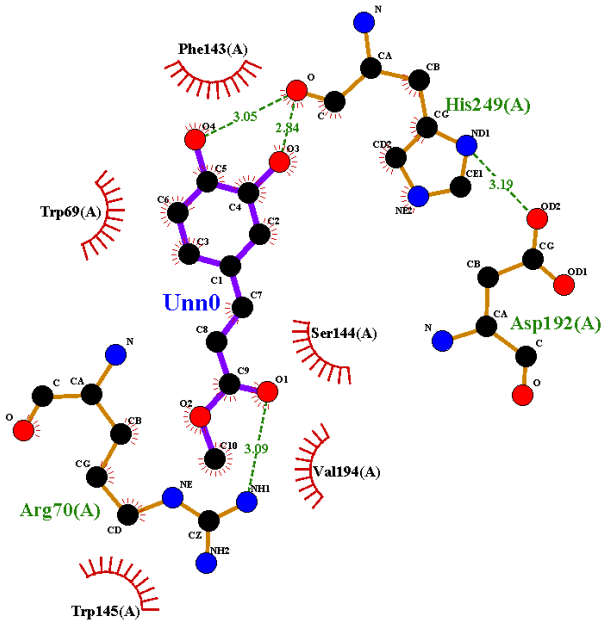
A-MFA



B-MCA


Arg68(A)
Trp67(A)
His247(A)
Val192(A)
Trp143(A)
Unn0
Ser142(A)
Leu193(A)
2.79
2.57
2.93

C-MCA


Phe143(A)
His249(A)
Trp69(A)
Unn0
Ser144(A)
Asp192(A)
Val194(A)
Arg70(A)
Trp145(A)
3.05
2.84
3.19
3.09

D-MCA

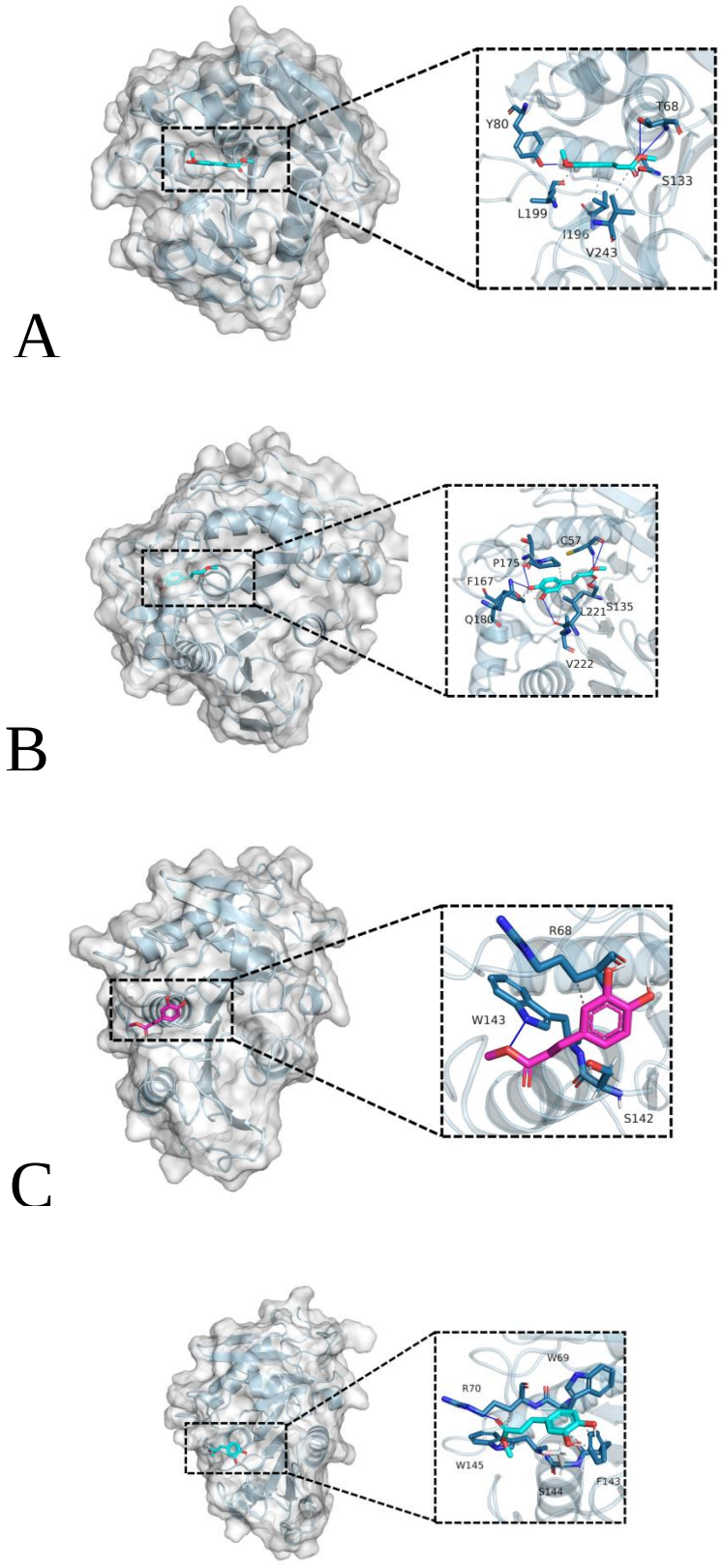


Figure 3.5 – Diagramming the Docking Results

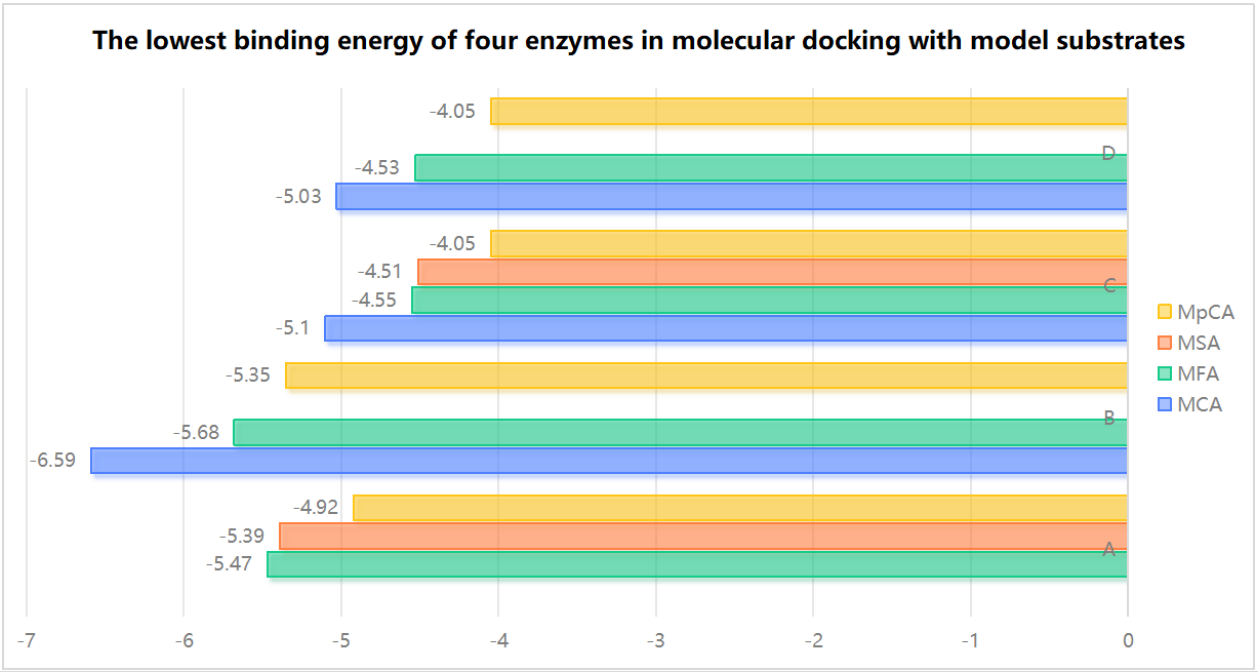
In type A enzymes, the substrate is stably bound to the active site through interactions with key amino acid residues such as Y80, I196, V243, T68 and S133. Substrate binding in type B enzymes, on the other hand, involves residues G57, F167, P175, Q180, S135, and V222. For type C enzymes, substrate interactions with residues such as R68, W143, and S142 were particularly significant. Whereas in type D enzymes, the substrate achieves binding mainly through interaction with residues such as W69, R70, W145 and F143.

3.3.2 Docking minimum binding energy analysis

In this study, we performed molecular docking experiments of four ferulic acid esterases (A, B, C, and D) with four model substrates (MCA, MFA, MSA, and MpCA) by using the Autodock tool with the aim of exploring the interactions between different enzymes and substrates and their stability.

The experimental results showed that the binding of enzyme A to MCA displayed the lowest binding energy (-6.59 kcal/mol), suggesting that this combination has the highest binding affinity and the most stable interaction. In contrast, the binding energy of enzyme D to MpCA was -4.05 kcal/mol, the highest of all combinations tested, indicating a relatively weak interaction.

In addition, a comparison of the binding energies of different enzymes for the same substrate revealed significant differences in the affinity of different enzymes for the same substrate. For example, for the MCA substrate, the binding energies of different enzymes ranged from -5.1 to -6.59 kcal/mol. This difference may stem from structural differences in the enzymes that affect how they interact with the substrate and how efficiently they do so.



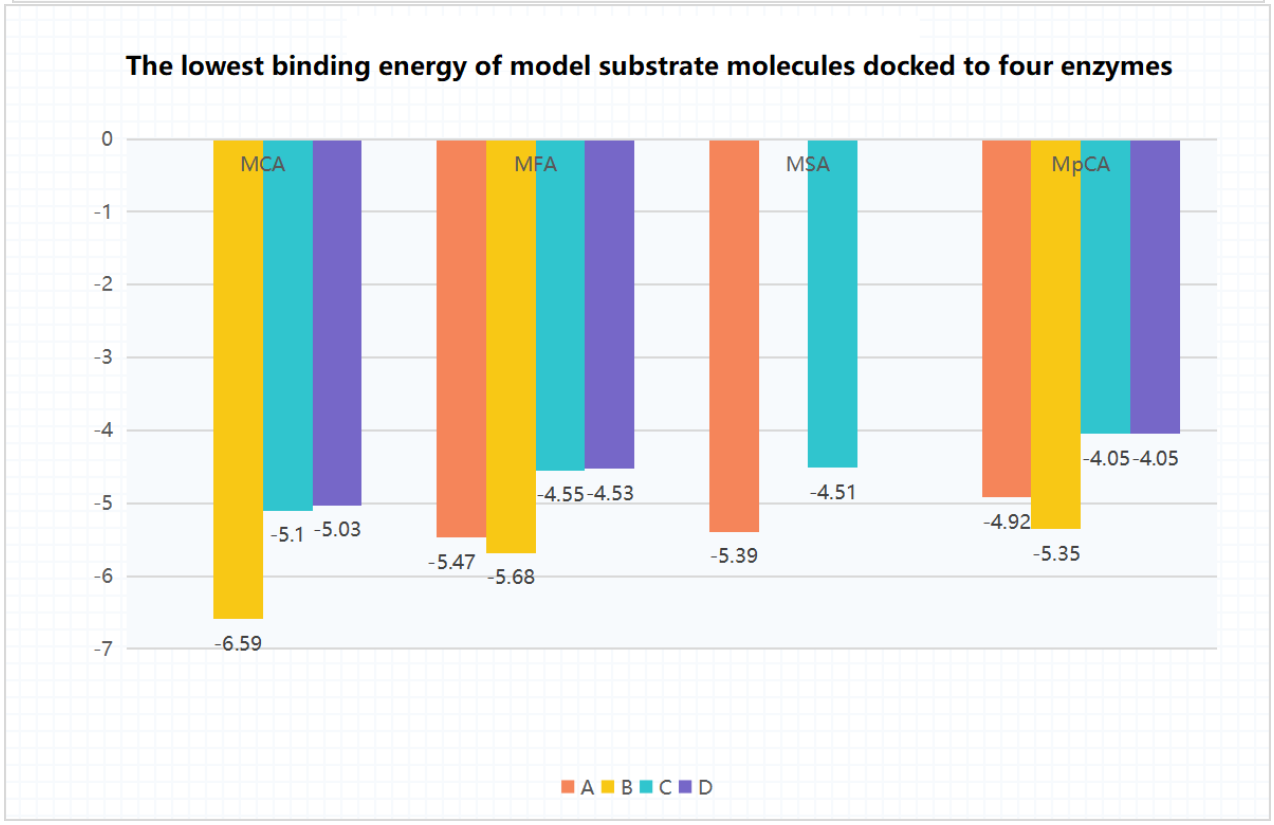
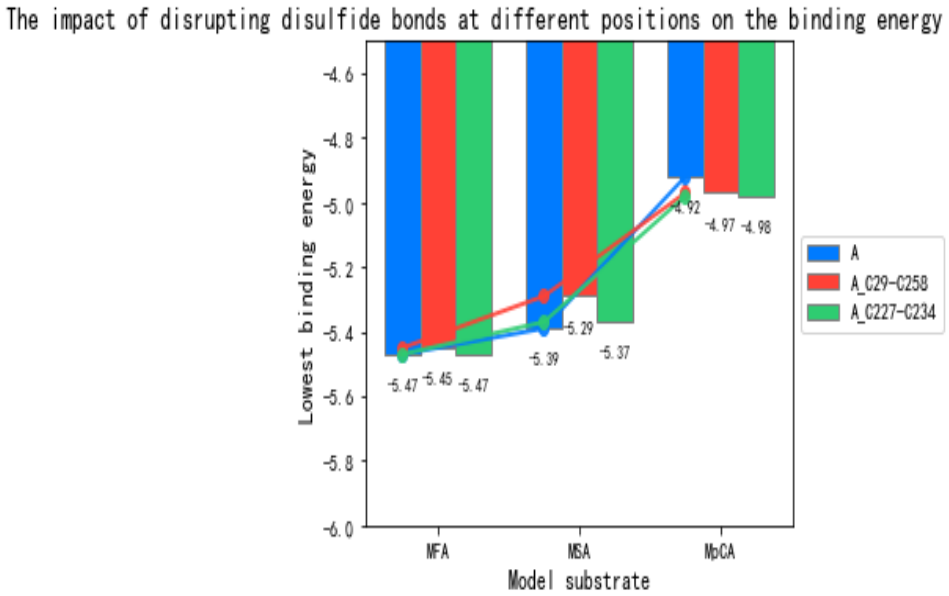


Figure 3.6 – Minimum binding energy results shown

3.4 Effect of disulfide bonds on binding energy

3.4.1 Change in minimum binding energy by knocking down/increasing disulfide bonds

This study shows that disulfide bonds play an important role in the structural stability and substrate binding ability of enzymes. For class A and B enzymes, knockdown of disulfide bonds usually leads to a decrease in binding energy, which may be attributed to the fact that the absence of disulfide bonds weakens the structural integrity of the enzyme, thus affecting its effective binding to substrates. However, for MpCA, which requires the most binding energy,, on the other hand, the required binding energy decreases, which suggests that a molecular dynamics study of this enzyme is warranted. For class C and D enzymes, the introduction of disulfide bonds decreased the binding energy required for binding to some substrates, which could be attributed to the enhanced structural stability of the enzyme by the addition of the disulfide bonds, thus improving its catalytic efficiency, while the binding energy required for some substrates increased instead, which requires further molecular dynamics verification. These results emphasize the importance of precise regulation of disulfide bonds in enzyme engineering to optimize the catalytic properties of enzymes.



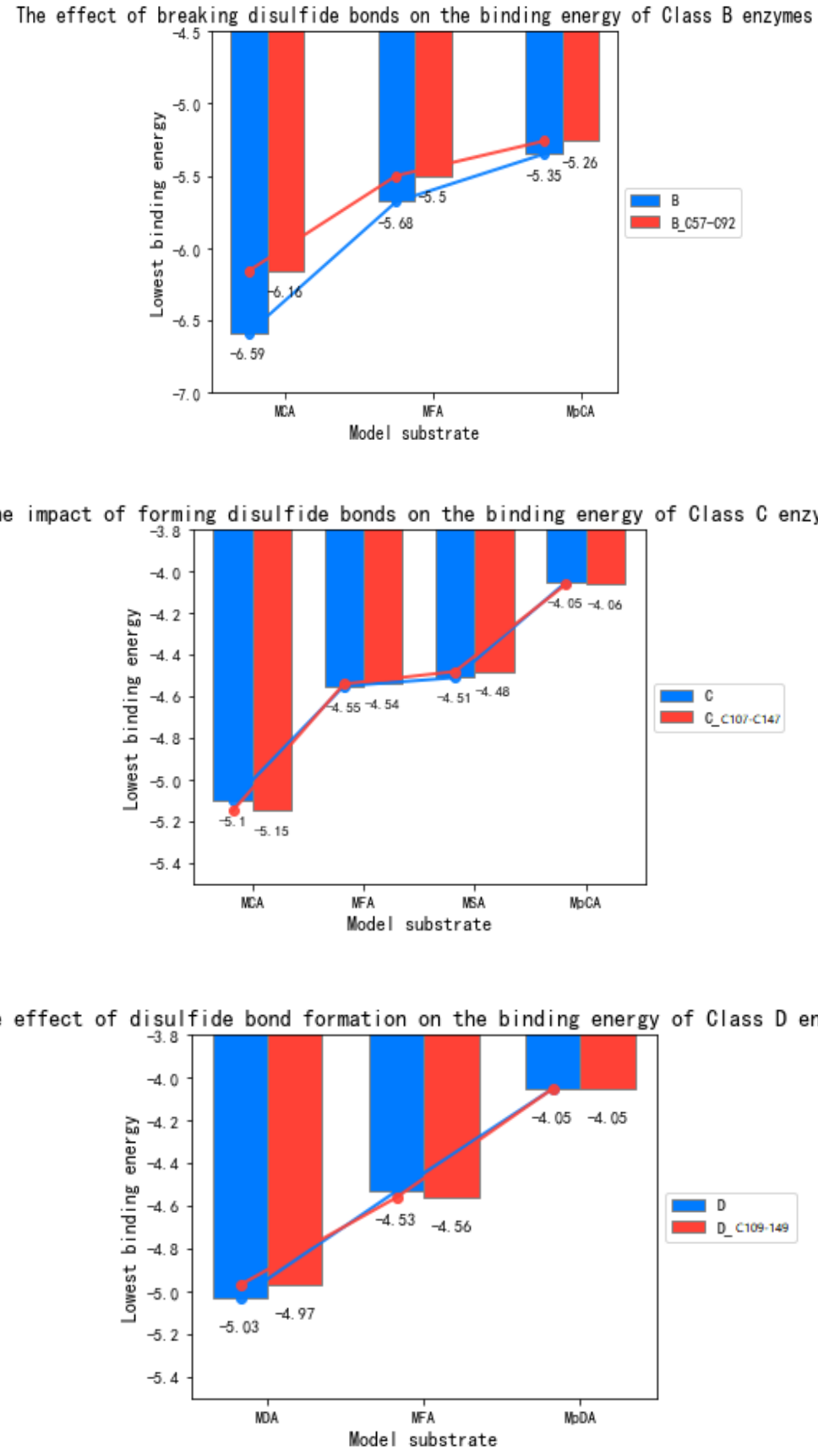


Figure 3.7 – Changes in enzyme-model substrate binding activity after knockdown/establishment of disulfide bonds

3.4.2 Molecular dynamics analysis

For the experimental design involving Class A enzymes, select the knockouts in the C29-C258 group that interact with MSA (Minimum Binding Affinity, MBA, increase of at least 0.1 kcal/mol) and MpCA substrates (MBA decrease of at least 0.05 kcal/mol) for molecular dynamics comparative studies. For Class B enzymes, choose the C57-C92 knockout group interacting with MCA as the experimental group. In Class C enzymes, select the disulfide bond addition group interacting with MCA (MBA decrease of at least 0.05 kcal/mol) and MSA (MBA increase of at least 0.03 kcal/mol) as the comparative experimental groups. For Class D enzymes, choose the disulfide bond addition group interacting with MCA (MBA increase of at least 0.06 kcal/mol) and MFA (MBA decrease of at least 0.03 kcal/mol) as the comparative experimental groups.

Table 3.1 – Molecular Dynamics Laboratory Group

	A	A_db-	Required binding energy		A	A_db-	Required binding energy
MSA	-5.39	-5.29	↑	MpCA	-4.92	-4.97	↓
	B	B_db-					
MCA	-6.59	-6.16	↑				
	C	C_db+			C	C_db+	
MCA	-5.1	-5.15	↓	MSA	-4.51	-4.48	↑
	D	D_db+			D	D_db+	
MCA	-5.03	-4.97	↑	MFA	-4.53	-4.56	↓

The interactions between class A feruloyl esterases and model substrates MSA and MpCA were investigated by analyzing the changes in RMSD (Root Mean Square

Deviation) and Rg (Radius of Gyration) before and after the disruption of disulfide bonds. This revealed the structural stability and dynamic changes of the enzymes.

As shown in Figure 3.8, when the disulfide bond C29-C258 was intact, the enzyme bound to MSA (A-MSA) exhibited significant fluctuations in RMSD, particularly between 5000 ps and 7500 ps, indicating substantial conformational changes upon MSA binding. This may be related to the higher binding activity of MSA. In contrast, the enzyme bound to MpCA (A-MpCA) had a lower and less variable RMSD, suggesting a more stable binding with MpCA. However, despite the slightly lower binding energy of A-MpCA compared to A-MSA, this also indicates that the structural flexibility of the enzyme may be positively correlated with its catalytic activity against model substrates.

A combined analysis of Figures 3.8 and 3.9 shows that after the disruption of the disulfide bond C29-C258, the RMSD curve of the enzyme bound to MSA (A-MSA-C29-C258) flattened out, and the Rg value decreased with reduced fluctuation, indicating that the overall structure of the enzyme became more compact and stable. However, this compactness may have led to an increase in binding energy, suggesting that the enzyme’s conformational changes were affected during the catalysis of MSA. For the enzyme bound to MpCA (A-MpCA-C29-C258), the Rg curve slightly increased with greater fluctuation, and the binding energy also decreased slightly, further supporting the positive correlation between enzyme conformational flexibility and its catalytic activity.

The structural stability and flexibility of class A feruloyl esterases show significant differences when binding to the model substrates MSA and MpCA. The disulfide bond C29-C258 plays a key role in maintaining the structural stability of the enzyme and regulating its binding activity with substrates. Before the disulfide bond was disrupted, the enzyme’s binding with MSA was accompanied by significant conformational changes, while the binding with MpCA was relatively stable. After the disruption, the overall structure of the enzyme became more compact, but this compactness may have had a relatively negative impact on the enzyme’s catalytic

activity. These findings provide important insights into understanding the structure-function relationship of class A feruloyl esterases.

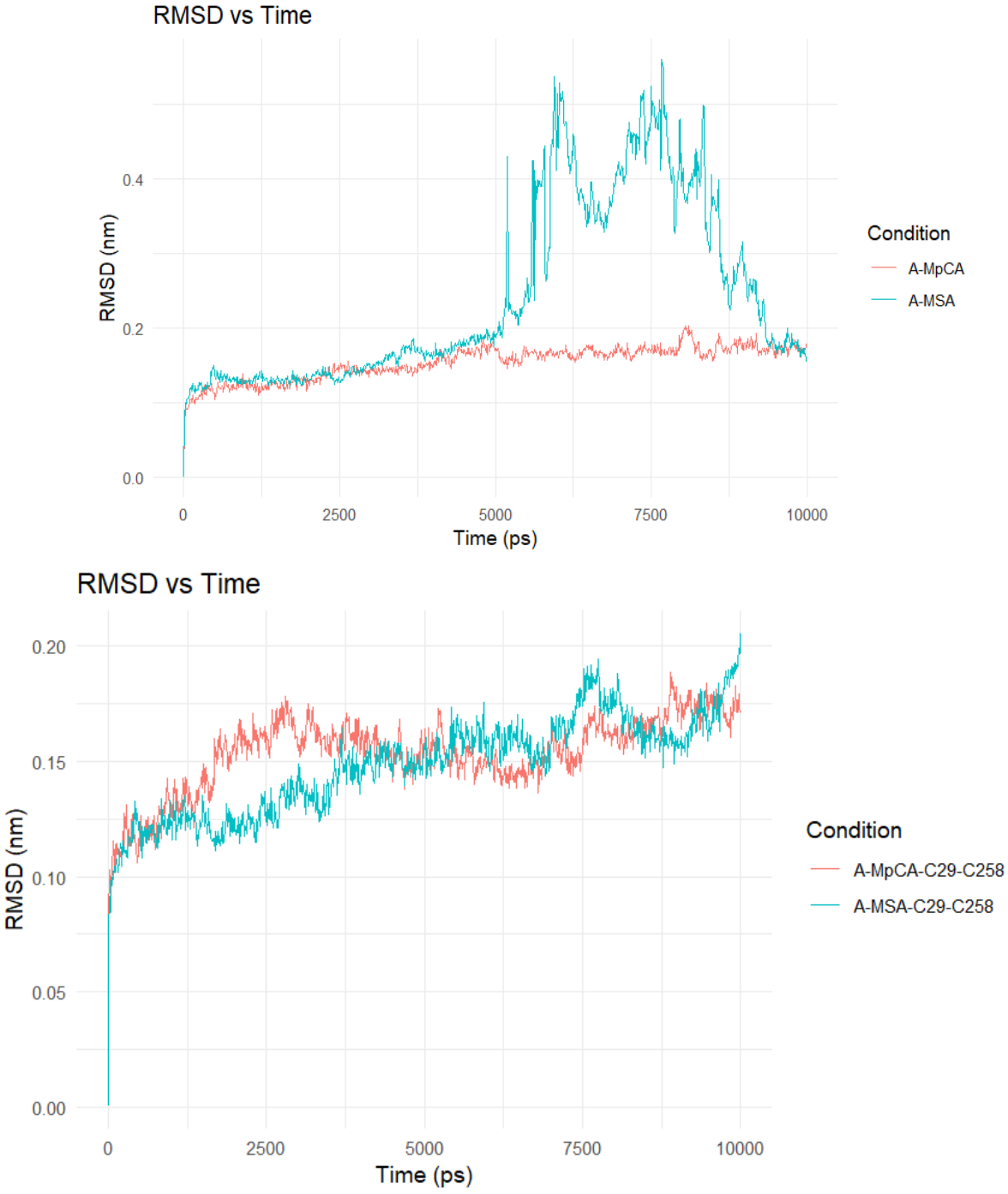


Figure 3.8 – RMSD analysis of class A enzyme with and without disulfide bond disruption, bound to model substrates MpCA and MSA.

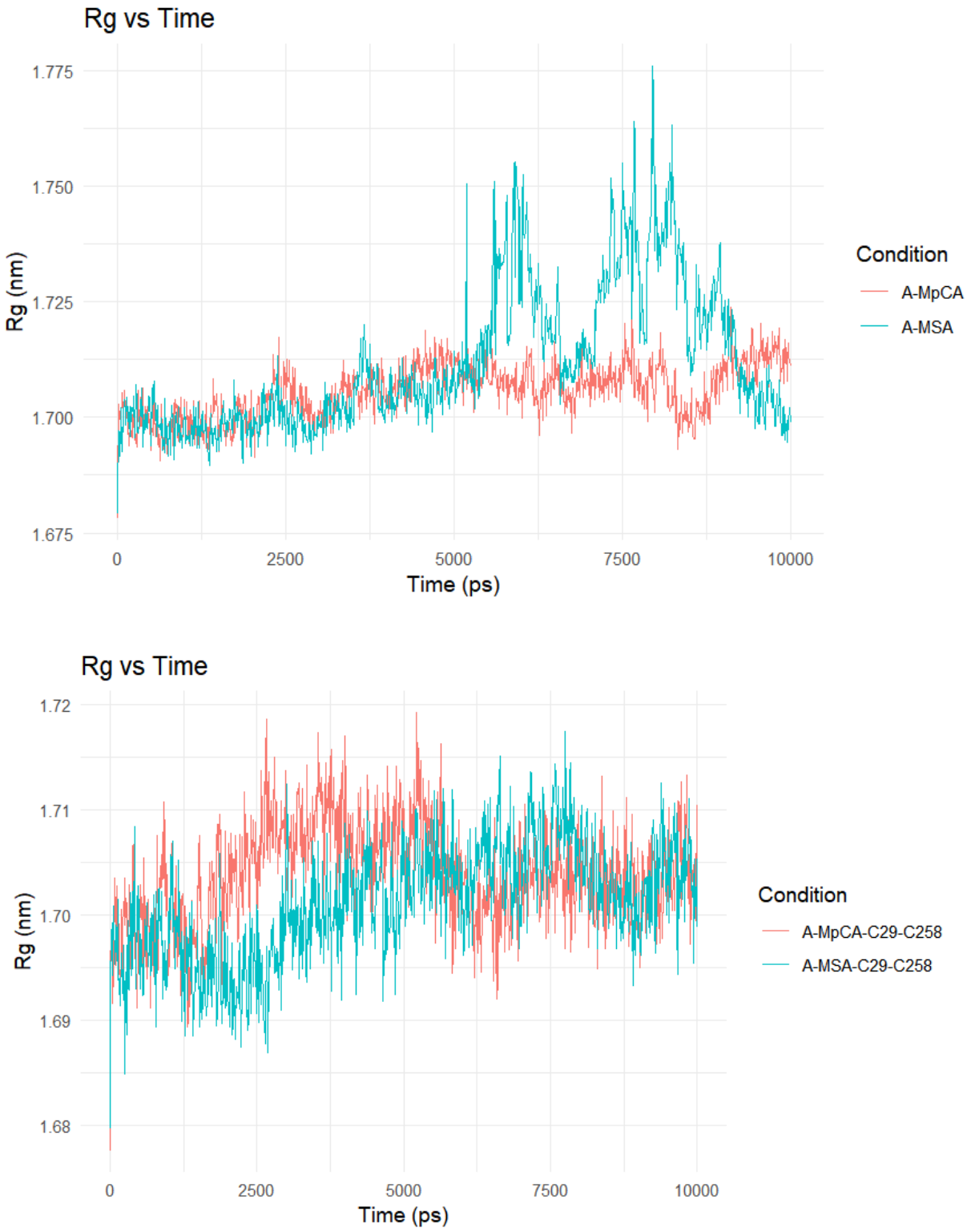


Figure 3.9 - Rg analysis of class A enzyme with and without disulfide bond disruption, bound to model substrates MpCA and MSA

To investigate the effect of disulfide bond disruption on the binding of class B feruloyl esterase (B-MCA) to the model substrate (MCA), molecular dynamics simulations were conducted, and the changes in RMSD and Rg over time were analyzed.

As shown in Figure 3.10, the RMSD values for the B-MCA system without disulfide bond disruption (green curve) remained at a relatively low level throughout the simulation, around 0.2 nm, indicating a structurally stable system. However, the RMSD values significantly increased for the B-MCA-C57-C92 system with disulfide bond disruption (blue curve), especially at the beginning of the simulation (0-5000 ps), before stabilizing but still remaining higher than the undisrupted system. This suggests that the disruption of disulfide bonds led to significant structural changes and increased instability in class B feruloyl esterase.

Figure 3.11 shows the changes in Rg over time for the two systems. The Rg value for the B-MCA system without disulfide bond disruption (green curve) remained relatively stable around 1.75 nm throughout the simulation. In contrast, the Rg value for the B-MCA-C57-C92 system with disulfide bond disruption (blue curve) gradually increased during the simulation, reaching approximately 1.83 nm. This further indicates that the disruption of disulfide bonds led to a looser and more extended protein structure.

The disruption of disulfide bonds significantly affected the structural stability and binding capacity of feruloyl esterase class B to the model substrate. These results provide a theoretical basis for further understanding the importance of disulfide bonds in the structure and function of enzymes.

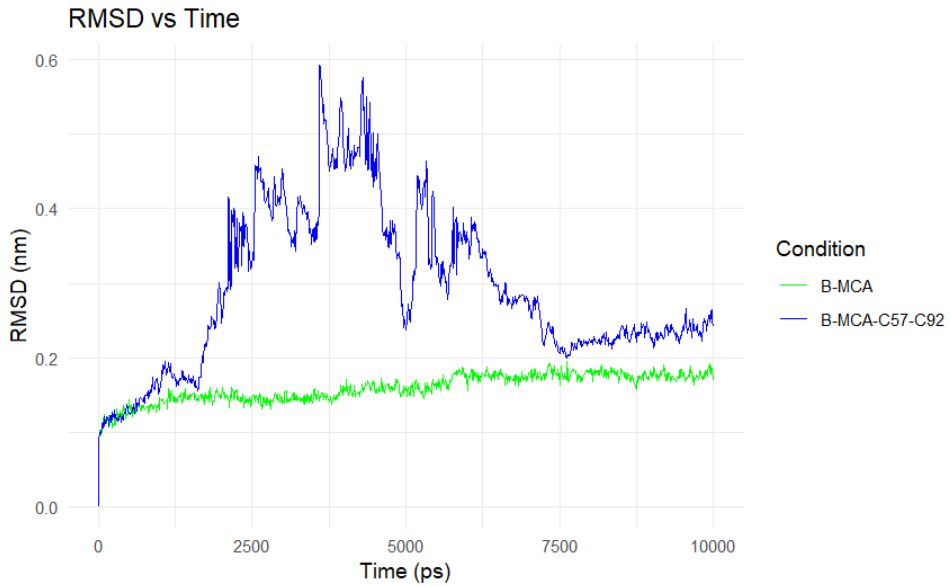


Figure 3.10 - RMSD analysis of class B enzyme with and without disulfide bond disruption, bound to the model substrate MCA

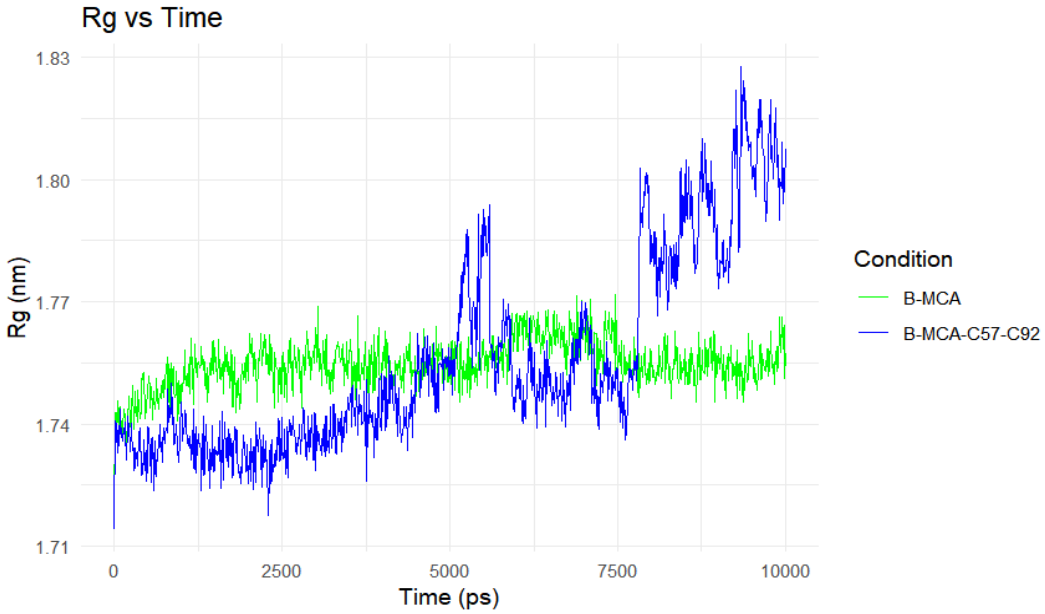


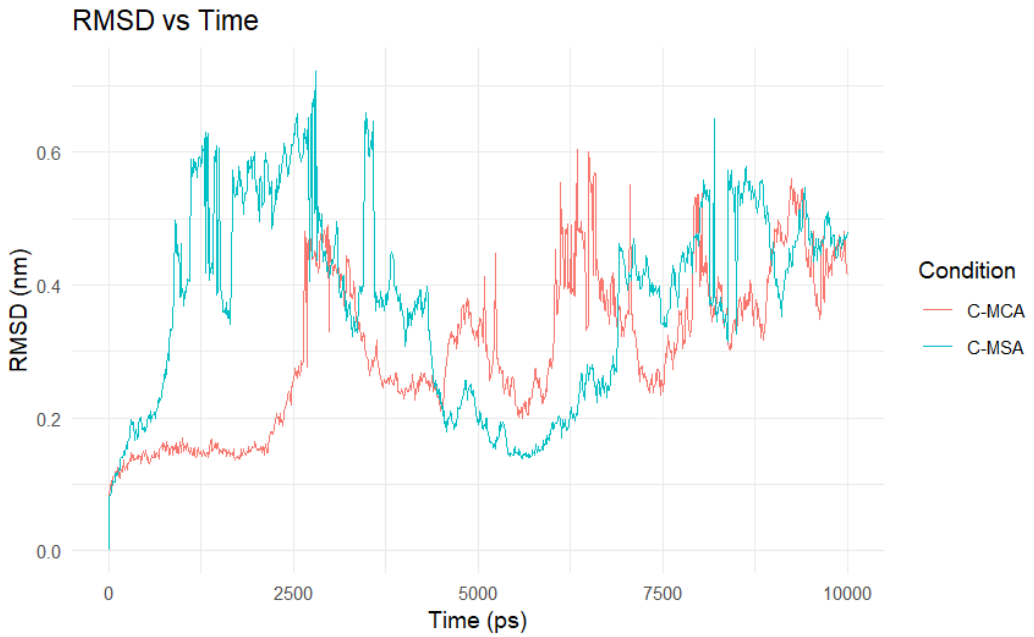
Figure 3.11 - Rg analysis of class B enzyme with and without disulfide bond disruption, bound to the model substrate MCA.

Figure 3.12 shows the RMSD changes of feruloyl esterase class C when binding to the model substrate MCA, with and without the addition of disulfide bonds. It can be observed that in the absence of additional disulfide bonds (C-MCA), the RMSD values fluctuated significantly throughout the simulation, especially around 2500 ps

and 7500 ps, where the RMSD values reached approximately 0.6 nm. However, after the addition of disulfide bonds (C-MCA-db+), the overall RMSD values decreased, and the range of fluctuation also narrowed, with a maximum value of about 0.4 nm. This indicates that the binding of the enzyme to the model substrate MCA became more stable after the addition of disulfide bonds.

Figure 3.13 presents the Rg changes of feruloyl esterase class C when binding to the model substrate MCA, with and without the addition of disulfide bonds. In the absence of additional disulfide bonds (C-MCA), the Rg values fluctuated greatly throughout the simulation, with a maximum value close to 1.825 nm. After the addition of disulfide bonds (C-MCA-db+), the fluctuation range of the Rg values significantly narrowed, with a maximum value of about 1.8 nm. This further suggests that the structure of the enzyme became more compact and stable after the addition of disulfide bonds.

The addition of disulfide bonds significantly improved the binding stability of feruloyl esterase class C to the model substrate MCA, as evidenced by the decrease in RMSD and Rg values and the increase in binding energy. These results indicate that the structural stability and substrate binding capacity of the enzyme can be effectively enhanced by increasing the number of disulfide bonds.



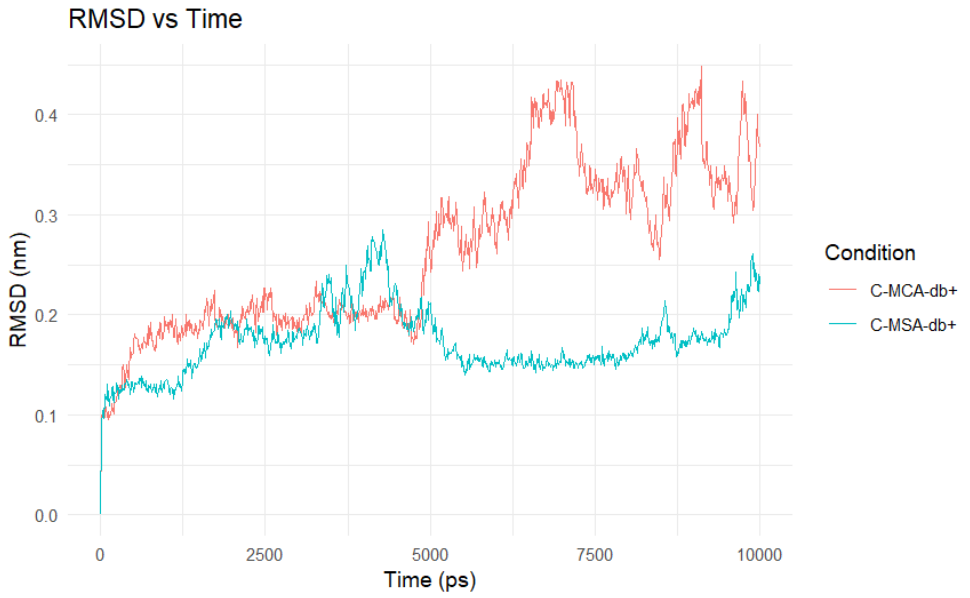
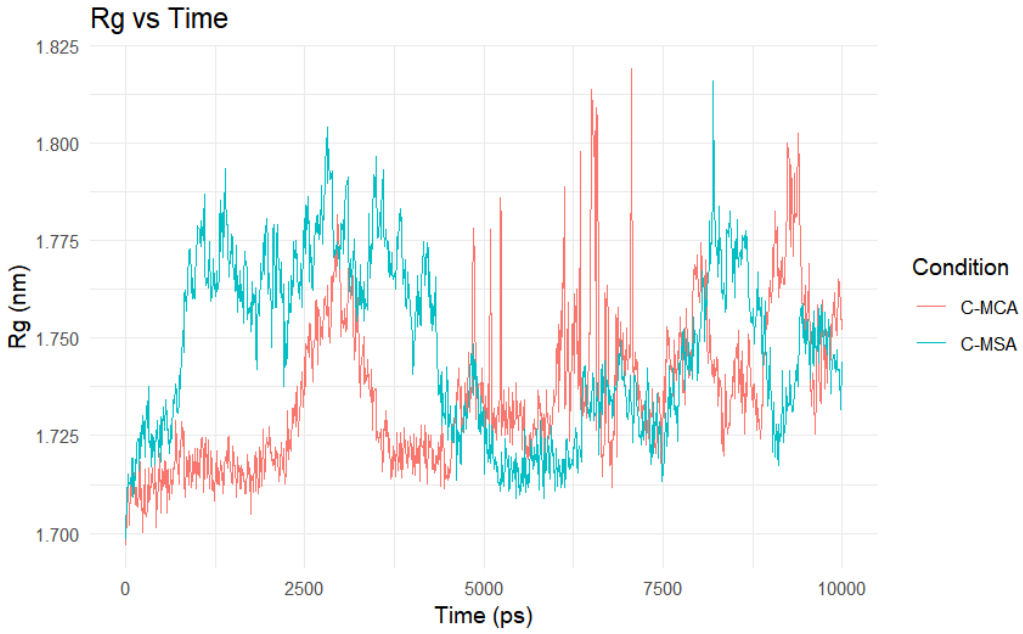


Figure 3.12 - RMSD analysis of class C enzyme with and without disulfide bond disruption, bound to the model substrate MCA.



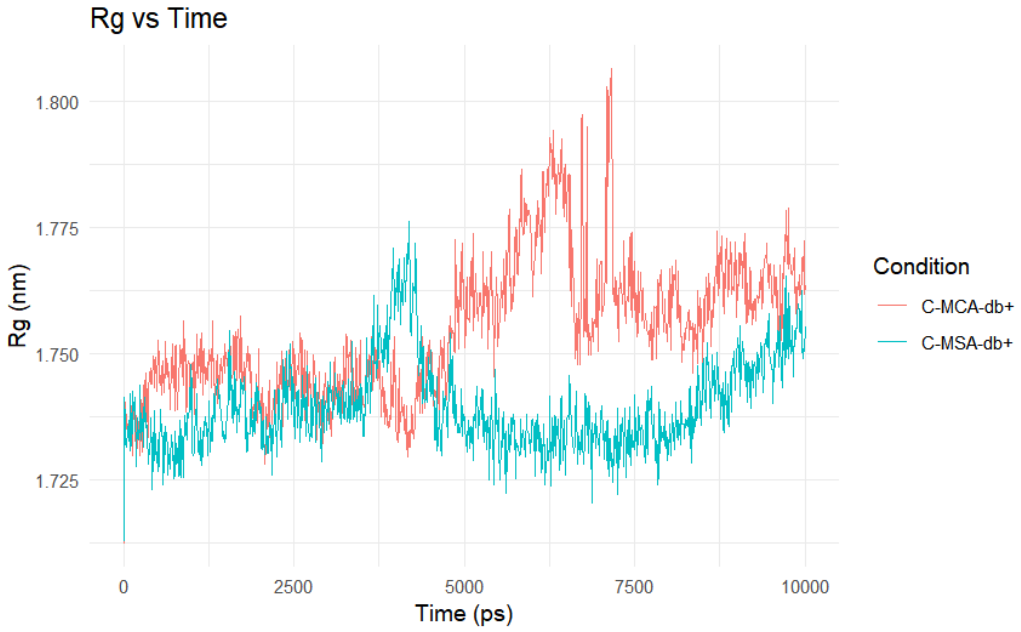


Figure 3.13 – Rg analysis of class C enzyme with and without disulfide bond disruption, bound to the model substrate MCA.

Figure 3.14 presents the RMSD changes in the molecular structure of feruloyl esterase class D due to the addition of disulfide bonds when binding to the standard substrates MCA and MFA. The results showed that after the introduction of additional disulfide bonds, the RMSD values of the enzyme when bound to MCA exhibited a more stable trend, with not only lower overall values but also a reduced fluctuation amplitude, demonstrating that the enzyme conformation gained higher stability due to the increased disulfide bonds. When bound to the model substrate MFA, the RMSD values also decreased after the addition of disulfide bonds, but with a greater fluctuation amplitude, indicating that the enzyme structure was stabilized to some extent but still exhibited some degree of variability.

Figure 3.15 shows the Rg changes of feruloyl esterase class D when binding to the model substrates MCA and MFA, with and without the addition of disulfide bonds. The results showed that after the addition of disulfide bonds, the Rg values of the enzyme when bound to MCA were relatively lower and stable throughout the simulation, indicating a more compact overall structure. When bound to the model

substrate MFA, the Rg values also decreased after the addition of disulfide bonds, but with greater fluctuations, indicating that the enzyme structure became more compact to some extent but still exhibited some structural looseness.

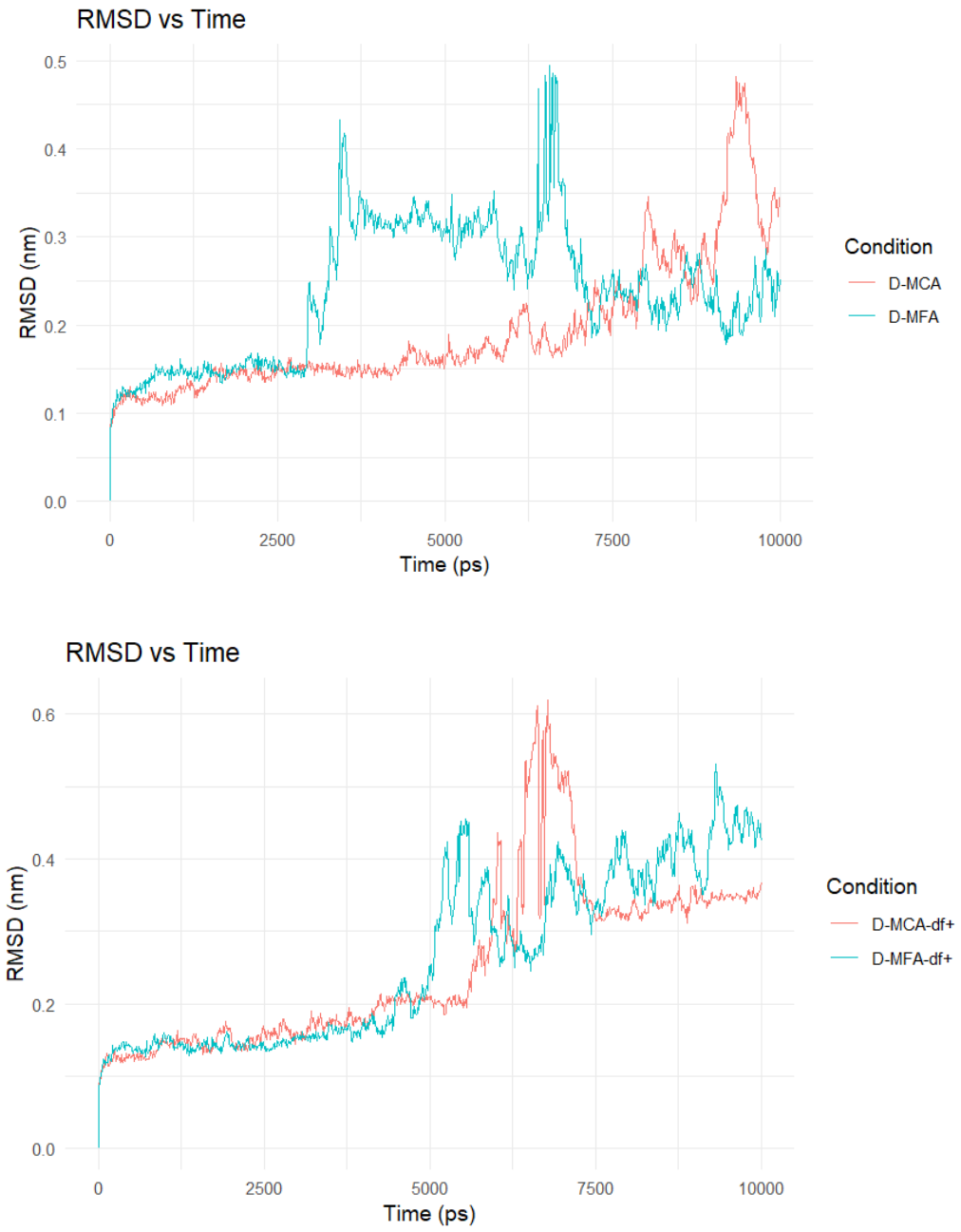


Figure 3.14 – RMSD analysis of class D enzyme with and without disulfide bond disruption, bound to the model substrate MCA.

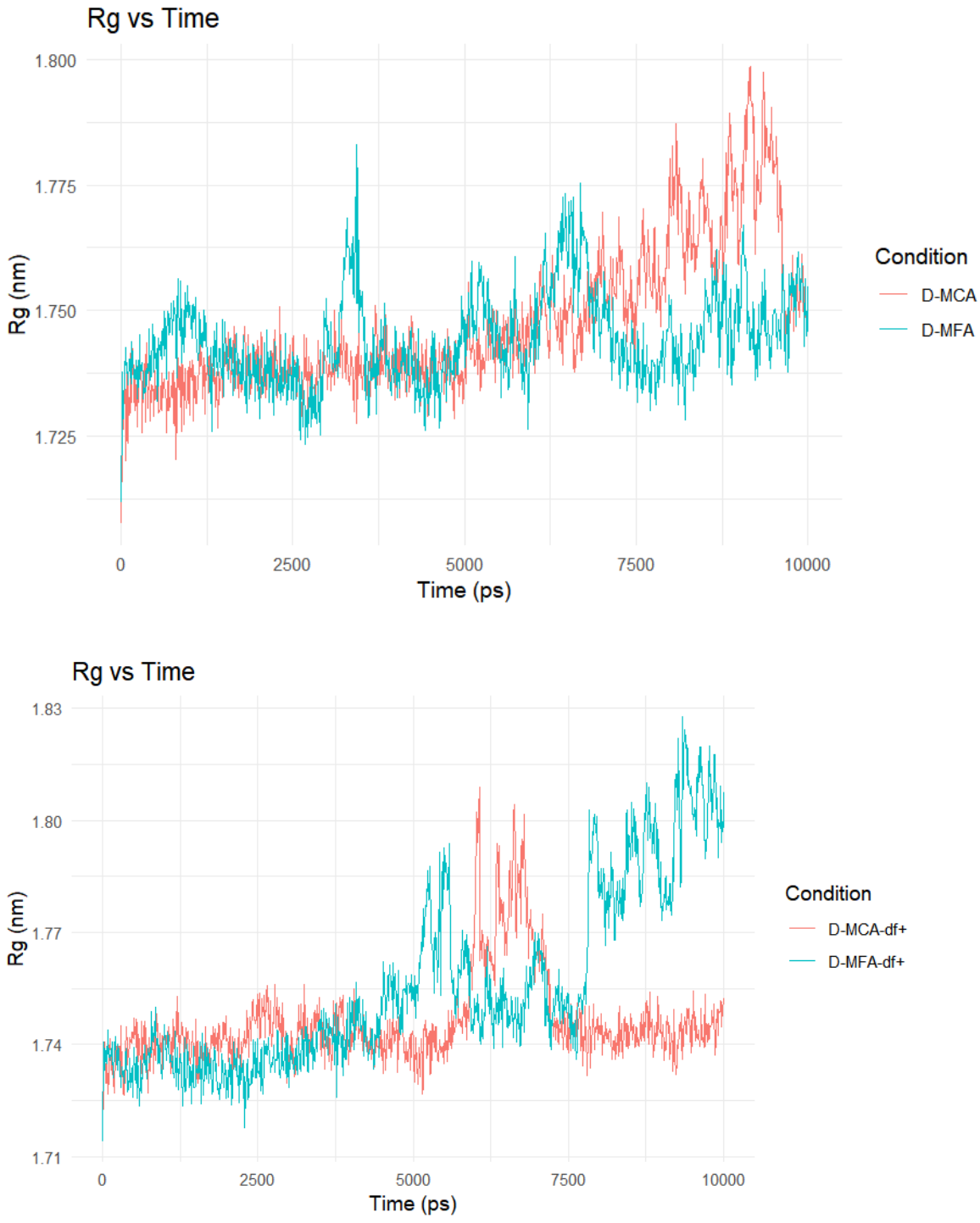


Figure 3.15 – Rg analysis of class D enzyme with and without disulfide bond disruption, bound to the model substrate MCA.

CONCLUSION

This study, through sequence analysis and homology modeling of four distinct types of feruloyl esterases, found that Enzyme A has a relatively compact structure, which may imply its adaptation for functioning within smaller organelles. In contrast, Enzyme B exhibits a more open structure, potentially facilitating interactions with larger polymers or substrates. This suggests that the structural features of different feruloyl esterases are closely associated with their functions and subcellular locations, providing insights into the structure-function relationships of enzymes.

To further investigate the impact of disulfide bonds on the structure and function of feruloyl esterases, site-directed mutagenesis was employed to disrupt and create disulfide bonds in the four enzyme types. The results showed that disruption of the C29-C258 disulfide bond in Enzyme A decreased its affinity for substrates but enhanced structural stability. This increased stability, however, inhibited the binding activity of Enzyme A to the model substrate MSA, with the lowest binding energy changing from -5.39 kcal/mol to -5.29 kcal/mol. It is postulated that the catalysis of this model substrate by Enzyme A requires relatively flexible conformational changes, which are inhibited by the disulfide bond disruption. However, this also results in a tighter binding to the model substrate MpCA with relatively lower activity, lowering the minimum activation energy required for binding from -4.92 kcal/mol to -4.97 kcal/mol. In Enzyme B, the disrupted disulfide bond is located near the active site within a cysteine loop, and its breakage significantly affects the enzyme's substrate binding capacity, causing a substantial increase in the lowest binding energy required for the model substrates, with the most pronounced effect on the most active substrate MCA, changing from -6.59 kcal/mol to -6.16 kcal/mol. Additionally, the introduction of disulfide bonds in Enzymes C and D enhanced their stability, but the effects on binding affinity were substrate-dependent. These findings indicate that disulfide bonds play a pivotal role in maintaining the structural stability and catalytic activity of feruloyl esterases, although their impact varies with the enzyme type and substrate.

Through systematic sequence analysis, homology modeling, and site-directed mutagenesis experiments, this study has elucidated the molecular mechanisms underlying the substrate specificity of four different types of feruloyl esterases. The results suggest that the structural features of enzymes, the positioning and number of disulfide bonds, and the nature of substrates collectively determine the catalytic activity and substrate specificity of enzymes. Different types of feruloyl esterases exhibit distinct responsiveness patterns, indicating that structural characteristics and regulatory mechanisms should be carefully considered in enzyme engineering endeavors, necessitating targeted strategies. This research offers a novel perspective on understanding the structure-function relationships of enzymes and holds exploratory significance for the fields of enzyme engineering and biocatalysis.

LIST OF REFERENCES

1. Crepin, V. F., Faulds, C. B., & Connerton, I. F. (2004). Functional classification of the microbial feruloyl esterases. Applied Microbiology and Biotechnology, 63(6), 647-652.

2. Mathew, S., & Abraham, T. E. (2004). Ferulic acid: an antioxidant found naturally in plant cell walls and feruloyl esterases involved in its release and their applications. Critical Reviews in Biotechnology, 24(2-3), 59-83.

3. Fazary, A. E., & Ju, Y. H. (2007). Feruloyl esterases as biotechnological tools: current and future perspectives. Acta Biochimica et Biophysica Sinica, 39(11), 811-828.

4. Topakas, E., Vafiadi, C., & Christakopoulos, P. (2007). Microbial production, characterization and applications of feruloyl esterases. Process Biochemistry, 42(4), 497-509.

5. Faulds, C. B. (2010). What can feruloyl esterases do for us?. Phytochemistry Reviews, 9(1), 121-132.

6. Dilokpimol, A., Mäkelä, M. R., Aguilar-Pontes, M. V., et al. (2016). Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. Biotechnology for biofuels, 9(1), 1-18.

7. Oliveira, D. M., Mota, T. R., Oliva, B., et al. (2019). Feruloyl esterases: Biocatalysts to overcome biomass recalcitrance and for the production of bioactive compounds. Bioresource technology, 278, 408-423.

8. Gupta, A., Sharma, P., Capalash, N., et al. (2019). Computational approaches for predicting substrate specificity of microbial carbohydrate active enzymes. Current Protein and Peptide Science, 20(3), 212-223.

9. Benoit, I., Navarro, D., Marnet, N., et al. (2006). Feruloyl esterases as a tool for the release of phenolic compounds from agro-industrial by-products. Carbohydrate Research, 341(11), 1820-1827.

10. Waterhouse, A., Bertoni, M., Bienert, S., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic acids research, 46(W1), W296-W303.

11. Xiang, Z. (2006). Advances in homology protein structure modeling. Current Protein and Peptide Science, 7(3), 217-227.

12. Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. Science, 294(5540), 93-96.

13. Ferreira, L. G., Dos Santos, R. N., Oliva, G., et al. (2015). Molecular docking and structure-based drug design strategies. Molecules, 20(7), 13384-421.

14. Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. Neuron, 99(6), 1129-1143.

15. Alonso, H., Bliznyuk, A. A., & Gready, J. E. (2006). Combining docking and molecular dynamic simulations in drug design. Medical Research Reviews, 26(5), 531-68.

16. Dilokpimol, A., Mäkelä, M. R., Cerullo, G., et al. (2018). Fungal feruloyl esterases: Functional validation of genome mining based enzyme discovery including uncharacterized subfamilies. New Biotechnology, 41, 9-14.

17. Tacias-Pascacio, V. G., Virgen-Ortíz, J. J., Jiménez-Pérez, M., et al. (2019). Evaluation of different lipase biocatalysts in the production of biodiesel from used cooking oil: Critical role of the immobilization support. Fuel, 235, 1544-1553.

18. Udatha, D. R. B. K., Mapelli, V., Panagiotou, G., et al. (2012). Common and distant structural characteristics of feruloyl esterase families from Aspergillus oryzae. PLoS One, 7(6), e39473.

19. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.

20. Laskowski, R. A., MacArthur, M. W., Moss, D. S., et al. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography, 26(2), 283-291.

21. Morris, G. M., Huey, R., Lindstrom, W., et al. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of Computational Chemistry, 30(16), 2785-2791.

22. Forli, S., Huey, R., Pique, M. E., et al. (2016). Computational protein–ligand docking and virtual drug screening with the AutoDock suite. Nature Protocols, 11(5), 905-919.

23. Holm, L. (2020). DALI and the persistence of protein shape. Protein Science, 29(1), 128-140.

24. Tan, K. P., Nguyen, T. B., Patel, S., et al. (2013). Depth: a web server to compute depth, cavity sizes, detect potential small-molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. Nucleic Acids Research, 41(W1), W314-W321.

25. Teufel, F., Almagro Armenteros, J. J., Johansen, A. R., et al. (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. Nature Biotechnology, 40(9), 1023-1025.

26. Kumar, S., Stecher, G., Li, M., et al. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution, 35(6), 1547-1549.

27. Saitou, N., & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution, 4(4), 406-425.

28. Liu, Y., Grimm, M., Dai, W. T., et al. (2020). CB-Dock2: a new method for the prediction of protein-ligand binding poses. Briefings in Bioinformatics, 21(5), 1777-1786.

29. Hermoso, J. A., Sanz-Aparicio, J., Molina, R., et al. (2004). The Crystal Structure of Feruloyl Esterase A from Aspergillus niger Suggests Evolutive Functional Convergence in Feruloyl Esterase Family. Journal of Molecular Biology, 338(3), 495-506.

30. Phienluphon, A., Kondo, K., Mikami, B., et al. (2023). Structural insights into the molecular mechanisms of substrate recognition and hydrolysis by feruloyl esterase from Aspergillus sydowii. International Journal of Biological Macromolecules, 253, 127188.

31. Vanommeslaeghe, K., Hatcher, E., Acharya, C., et al. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry, 31(4), 671-690.

32. Craig, D. B., & Dombkowski, A. A. (2013). Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. BMC Bioinformatics, 14(1), 1-7.