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 **Bioinformatics Analysis of the Two-Component Signal Transduction System NtrYX in *Cereibacter azotoformans***

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

Completed: student of group BEBT-20
Shi GUOTAO

Scientific supervisor
Ihor HRETSKYI, 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
Shi Guotao

1. Thesis topic **Bioinformatics Analysis of the Two-Component Signal Transduction System NtrYX in *Cereibacter azotoformans***
Scientific supervisor Ihor Hretskyi, 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	Section 1 Literature review	From 06 April 2024 to 20 April 2024	
3	Section 2 Object, purpose, and methods of the research	From 21 April 2024 to 30 April 2024	
4	Section 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 12 May 2024 to 24 May 2024	
7	Submission of qualification work to the supervisor for feedback (14 days before the defense)	From 24 May 2024 to 10 June 2024	
8	Submission of bachelor's thesis to the department for review (12 days before the defense)	13 June 2024	
9	Checking the bachelor's thesis for signs of plagiarism (10 days before the defense)	15 June 2024	
10	Submission of bachelor's thesis for approval by the head of the department (from 7 days before the defense)	17 June 2024	

I am familiar with the task:

Student _____ Shi GUOTAO

Scientific supervisor _____ Ihor HRETSKYI

SUMMARY

Shi Guotao. Bioinformatics analysis of the NtrYX Two-component signal transduction system in *Cereibacter azotoformans*. – Manuscript.

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

Cereibacter azotoformans is a facultative anaerobic bacterium with important nitrogen fixation ability, and its NtrYX two-component signal transduction system plays a key role in regulating nitrogenase activity. The NtrYX two-component signal transduction system is an important regulatory mechanism within bacteria. The system perceives and responds to environmental signals through the interaction between HK (histidine kinase) and RR (response regulatory protein). However, there is currently a lack of research on the structure and function of the *C. azotoformans* NtrYX two-component system in China.

This study used bioinformatics methods to deeply analyze the structure and function of the *C. azotoformans* NtrYX system in China. Through sequence retrieval and domain analysis, we identified the key structural domains and their functions of NtrY and NtrX proteins, NtrY includes HAMP PAS, PAC, HisKA, HATPase_c, NtrX includes REC and AAA domains. The COACH tool was used to predict potential signal molecules in the extracellular signal sensing region of NtrY, and the interaction between the active region of NtrY's extracellular signal sensing region and small molecules was analyzed using and LigPlot+. Among them, Arg215 and Arg198 have strong hydrogen bonding interactions with small molecules. The molecular docking between NtrY and ATP indicates that the stimulation of external signals strengthens the alpha helix interaction of the HisKA domain, making it active and hydrogen bonding stronger, leading to changes in the HisKA structure and providing spatial distance for phosphorylation. By using the COACH server to predict the interaction between NtrX and ADP, it was found that the residue Ser173 Gly174, Val354, and Arg355 have strong hydrogen bonding interactions with ADP. In addition, the simulated interaction between NtrY and NtrX proteins indicates that, the

binding between the PAS region and the REC region is crucial for the signal transduction chain. This study analyzed the structural information of each key link in the NtrYX two-component signal transduction system using bioinformatics methods, and preliminarily revealed the NtrYX two-component signal transduction mechanism. The conclusion of the study provides theoretical support for protein engineering modification to improve the signal response level of the NtrYX two-component system, laying a foundation for further research on the mechanism of the *C. azotoformans* NtrYX two-component system.

Keywords: Cereibacter azotoformans; NtrYX two-component system; homology modeling; molecular docking; protein-protein interaction.

TABLE OF CONTENTS

INTRODUCTION

Cereibacter azotoformans is a facultative anaerobic bacterium with nitrogen-fixing capabilities that can play a significant role in agricultural production and environmental protection. The NtrYX two-component signal transduction system is widely present in bacteria and plays a key role in bacterial adaptation to environmental changes and regulation of physiological processes.

This study aims to use bioinformatics methods to conduct an in-depth analysis of the structure and function of the NtrYX system in *C. azotoformans*, to understand its role in the regulation of nitrogenase activity, as well as its potential environmental adaptability. The research also intends to provide new ideas and methods for future biotechnological applications.

This study utilized bioinformatics methods for research, employing various bioinformatics tools and techniques. Specifically, the BLAST tool was used to retrieve protein sequences similar to the target protein. The SMART tool was applied to analyze the protein domains. Homology modeling was adopted to construct the 3D structures of NtrY and NtrX proteins. The CB-Dock2 was used to dock the NtrY dimer with ATP molecules, and LigPlot+ software was utilized to analyze the interaction forces. Additionally, ClusPro was employed to predict the three-dimensional structure of protein-protein complexes.

A thorough understanding of the structure and function of the NtrYX two-component signal transduction system is of great importance for advancing research on *C. azotoformans*, revealing its nitrogen-fixing mechanisms, and exploring its application potential. Moreover, the study contributes to the development of two-component systems and bioinformatics, offering new perspectives and methodologies for the biotechnology field, and may also provide potential targets for the development of antimicrobial drugs. In summary, this paper conducts a comprehensive analysis of the NtrYX two-component signal transduction system in *Cereibacter azotoformans* using bioinformatics methods, aiming to reveal its

functions and regulatory mechanisms in bacterial physiological processes, and to provide a scientific basis for research and application in related fields.

The relevance of the topic is Bioinformatics analysis of the NtrYX Two-component signal transduction system.

The purpose of the study is the To provide a theoretical basis for further research on the physiological, biochemical, and ecological functions of the *Cereibacter azotoformans* strain, develop sustainable agriculture, protect the ecological environment, and solve environmental pollution issues.

The objectives of the study The analysis of NtrY and NtrX proteins through bioinformatics methods provides molecular insights into understanding the regulatory mechanism of the NtrYX two-component system in *C. azotoformans*.

The object of the study *Cereibacter azotoformans*.

The subject of the study The NtrYX Two-component signal transduction system.

Research methods Bioinformatics analysis methods.

The scientific novelty Bioinformatics analysis methods.

The practical significance of the results obtained is help us gain a deeper understanding of its transduction mechanism.

CHAPTER 1
LITERATURE REVIEW

1.1 Introduction to *Cereibacter azotoformans*

Cereibacter azotoformans is a Gram-negative bacterium belonging to the class *Alphaproteobacteria* [1]. This microorganism was first isolated from Japanese rice field soil in 2011 and was identified by scientists as a new genus and species [2]2.

C. azotoformans is a facultative anaerobe with a cell size approximately in the range of 0.6-0.8 μm × 1.5-2.0 μm [1]. This strain has a wide range of metabolic capabilities and can use a variety of carbon sources such as glucose, mannitol, and xylose for growth, and can use a variety of compounds such as ammonium salts, nitrates, nitrites, and amino acids as nitrogen sources [2, 3]. More significantly, *C. azotoformans* has nitrogenase activity and can fix nitrogen (N_2) in the atmosphere into ammonia (NH3), providing necessary nitrogen sources for plants [1-4]. This property makes *C. azotoformans* have great potential in agricultural production, as it can be used as a biofertilizer to improve soil fertility and reduce the need for chemical fertilizers.

In addition, the study also found that *C. azotoformans* can perform denitrification under anoxic conditions, reducing nitrate (NO_3^-) to nitrogen (N_2) and producing a certain amount of nitrous oxide (N_2O) [5]. This process is not only an important part of the soil nitrogen cycle, but may also have an impact on farmland greenhouse gas emissions [6]. Therefore, research on *C. azotoformans* can help us better understand and manage nitrogen cycling and greenhouse gas emissions from farmland. It is worth noting that *C. azotoformans* can also degrade some aromatic compounds that are difficult to degrade, such as benzoic acid and p-hydroxybenzoic acid [7, 8]. This property makes *C. azotoformans* have good application prospects in the field of environmental remediation, for example, it can be used to treat sewage or soil containing these harmful compounds.

C. azotoformans as a newly discovered microbial resource, it has broad application prospects in agricultural production, environmental protection, pollution

remediation and other fields. In-depth study of the physiological, biochemical characteristics and ecological functions of this strain is of great significance for us to develop sustainable agriculture, protect the ecological environment, and solve environmental pollution problems.

1.2 Introduction to Two-Component signal transduction systems

Two-Component signal transduction systems (TCS) is a widespread signaling mechanism within bacteria, enabling bacteria to detect internal and environmental signals, and plays a key role in bacteria adapting to environmental changes and regulating physiological processes. In addition, TCS is also found in some eukaryotes, but is significantly missing from the genomes of animals [9]. A typical two-component system consists ofhistidine histidine kinase (HK) and Response regulator (RR) consists of two parts [10]. NtrYX two-component system as one of the representatives, it consists of two proteins-NtrY and NtrX. NtrY is a histidine kinase protein, and NtrX is a response regulatory protein. The NtrYX two-componentsystem is present in many Proteobacteria, including many members of α-proteobacteria and β-proteobacteria, and is involved in a variety of cellular processes. The NtrYX two-component system has been widely studied in the bacterial domain and is often associated with the regulation of nitrogen metabolism, especially with processes such as nitrogen fixation, denitrification, and nitrate assimilation [11]. The NtrYX operon of prokaryotic cells encodes the sensor histidine kinase NtrY and the response regulator NtrX, which mediates the response of a variety of metabolic processes and is mainly involved in the regulation of nitrogen metabolism [11, 12], plays a central role in bacterial growth and metabolic balance. Studying this system will help us gain a deeper understanding of the molecular mechanisms of symbiotic nitrogen fixation. Recently, it was found that the NtrYX two-component system of *Rhodobacter sphaeroides* plays an important role in regulating the cell envelope, including the biosynthesis/modification of peptidoglycan and the cell division process [11]. Research on the NtrYX two-component system can help us reveal how this system

regulates bacterial physiological processes and the impact of these regulatory processes on bacterial survival and adaptation to the environment.

1.3 Structure and conduction mechanism of two-component signal transduction system

1.3.1 Basic structure of two-component signal transduction system

Two-component signaling systems (TCSs) are signal transduction pathways that enable bacteria to detect internal and environmental signals and link these signals to appropriate genetic and biochemical programs to adapt and survive [12]. A typical two-component system consists of two parts: histidine kinase (HK) and response regulator (RR), HK is mostly Transmembrane proteins, and RR is generally located in the cytoplasm [13]. TCS relies on phosphorylation transfer reactions as a means of transmitting information [14]. HK and RR play key roles in the phosphorylation transfer process of TCS. Under the stimulation of external environmental signals, the His site of HK undergoes autophosphorylation, and then transfers its phosphate group to the Asp site of the paired RR, causing changes in protein conformation [5]. Its autophosphorylation process includes: (a) spontaneous hydrolysis; (b) specific phosphatase; (c) in most cases, the phosphatase action of homologous HK in the kinase-off state [14]. The DNA-binding ability of phosphorylated RR is significantly increased, thereby activating or inhibiting the transcription of downstream target genes [13]. As the terminal of the two-component system, the phosphorylation level of RR is strictly controlled by the active state of HK, maintaining a dynamic balance of phosphorylation and dephosphorylation in the cell [13]. Histidine kinases (HKs) may also act as phosphatases to dephosphorylate phosphorylated response regulator proteins (RRs) [15]. TCSs that mediate host interactions in pathogens and symbionts are generally conserved in related free-living organisms and are capable of responding to similar physicochemical signals present in the environment [16].

There are four common molecular structures of the TCS pathway: PathwayI: Typical HK: RR pathway, involving transmembrane HK. Pathway II: Phosphate relay pathway with intermediate REC and HPt proteins. Pathway III: Complex HK

structures with longer transport modules. Pathway IV: involves soluble HK [14]. While typical TCSs rely on a single HK and RR pair, many systems incorporate other proteins, such as activators or inhibitors, to form more complex signaling networks [16].

Most TCSs are considered isolated systems, but in many bacteria, HK sand Cross-regulation between RRs may integrate multiple environmental signals. Even though HK and RR pairs are well isolated, TCSs can also interact at the transcriptional level. For example, one TCS may regulate the expression of other TCS genes, or multiple TCSs may affect the transcription of the same downstream gene [16].

1.3.2 Structure and function of HK and RR

Histidine kinases (HKs) are usually homodimeric proteins, and multiple different HKs are expressed in specific cells. HKs are divided into three major categories based on their modular design, namely I–III categories. The vast majority of HKs belong to class I, which consist of three structural/functional domains: (a) sensing module (SEN), (b) transport/transduction module (TRA), and (c) catalytic module (CAT) [14]. HKs often contain a TRA module connecting the sensing module and the catalytic module, which is composed of two or more helical transmembrane segments (in transmembrane HKs) and one or more signaling domains, such as HAMP (Histidine kinase, adenylate cyclase, methyl-accepting chemotactic protein and phosphatase), PAS (contained in Per-Arnt-Sim protein) or GAF (cGMP-specific phosphodiesterase, adenylate cyclase and FhlA) etc. [17].

CAT modules form a minimal structural core shared by all I classes HKs consisting of two Domain composition: dimerization and histidine phosphotransfer domains (DHp) and ABD. DHp is elongated and formed by a two-α-helical hairpin that drives homodimerization. Once dimers are formed, DHp forms an antiparallel four-helix bundle via the N-terminus of theα1helix Extend into two parallel helical structures. The phosphorylatable histidine is located within a conserved sequence motif (H box) on α1, approximately in the middle of the DHp [18,19]. ABD shows

aβfold: a globularα/βdomain at its core It consists of three helices, with fiveβstrands opposite one side. This fold is present in the GHKL protein superfamily (DNA gyrase, Hsp90, histidine kinase and MutL), all of which exhibit slow ATPase activity [20]. ABDs bind ATP-Mg+2, involving the definition of conserved motifs (N-,G1-,F-,G2- andG3-box) amino acids. The loop containing the F box serves as the true ATP lid (ATP-lid), forming the nucleotide-binding pocket and showing highly variable lengths in different families [14].

RRs always contain a REC domain within the conserved α/β Rossmann-like contains phosphorylatable aspartate residues within the fold. RRs can contain only RECs or other domains, mainly DNA-binding domains, but also ligand-binding domains, protein-binding domains, enzyme functions, etc [21]. Activation of RRs is related to their phosphorylation, and then P∼RRs participate in output effector responses (transcriptional regulation, allosteric regulation of chaperone activity, etc.) [14]. Seven families based on REC have currently been identified in Pfam, but one of these families includes known RR, the vast majority of sequences(Response_reg PF00072). On the other hand, analysis of output effector domains identified dozens of RR families, revealing extensive functional diversity21.

1.4 Function of NtrYX two-component signal transduction system

1.4.1 Functions of NtrYX two-component signal transduction system in different bacteria

Two-component signal transductionsy stems are ubiquitous in bacteria and also present in fungi and plants. So far, almost all studies on the NtrYX two-component system have been conducted in bacteria in which the NtrBC gene is located upstream of the NtrYX gene [22]. The NtrYX two-component system was originally discovered in *Azorhizobium caulinodans* via transposon mutagenesis, when it was described as a NtrBC-related TCS with a role in controlling nitrogen fixation [23]. In the growth, development and metabolic processes of bacteria, NtrYX two-component system plays an important role. In different bacteria, the role of the NtrYX two-component system in signal transduction is different.

Brucella spp is a facultative intracellular Gram-negative bacterium belonging to the genusα-2-Proteobacteria. They are pathogenic to many mammals, including humans, causing a disease called brucellosis [22]. In *Brucella* genus NtrYX two-component system as a redox sensor, with oxidation Reduction induction is related to denitrification and can regulate the expression of nitrogen respiratory enzymes. Studies have shown that the activity of HK NtrY is responsive to low oxygen tension. NtrY mutants activate nitrate reductase (narGHIJK), nitrite reductase (nirKV), nitric oxide under aerobic or microaerobic conditions. The expression of both reductase (norBCDEFQ) and nitrous oxide reductase (nosDFLRKYZ) gene clusters is down-regulated, and deletion of NtrYX reduces the expression of bioenergetic enzymes critical for survival in hypoxic environments [22, 24-25]. When *Brucella* is grown under aerobic conditions, NtrYis in its inactive, oxidized form. When oxygen tension is reduced, NtrYis destroyed by unknown cellular reducing agents such as the reduced quinone pool produced during anaerobic respiration or NADH) is reduced and autophosphorylation is activated. Then, the phosphorylated reduced NtrY transmits the signal to its cognate response regulator NtrX, which in turn interacts with the DNA promoter to activate genes involved in adaptation and detoxification of low oxygen tension, such as denitrification enzymes [22].

Paracoccus denitrificansis a α-Proteobacterium capable of denitrification under aerobic conditions. whereas the NtrYX two-component system is essential for paracoccus denitrificans to maintain cellular iron homeostasis and complete denitrification under iron-limiting conditions. Results indicate that the NtrYX two-component system is an iron-responsive transcriptional regulator that constitutes an optimal cellular iron homeostasis and diverse uses of nitrate (as a nitrogen source and energy source) The key regulatory elements [3]. The role of the NtrYX two-component system is particularly important under anaerobic and denitrifying conditions if the iron concentration in the environment fluctuates, especially when iron availability is very limited [11]. NtrYX two-componentsystem in Paracoccus denitrificans acts through its proteome, low intracellular iron content, and NtrY high concentrations of nitrous oxide released by the mutant were highlighted. In response

to iron limitation, the NtrYX two-component system not only controls iron homeostasis but also participates in the complete denitrification process. This function may also exist in the NtrYX two-component system in other microorganisms [11].

Neisseria gonorrhoeae is a Gram-negative diplococcus that is the causative agent of gonorrhea. The deletion of the NtrX gene in *Neisseria gonorrhoeae* affects the expression of genes involved in aerobic and anaerobic respiration. Characterization analysis of NtrX mutants in *Neisseria gonorrhoeae* shows that this response regulator regulates the expression of respiratory enzymes. plays a role in the bacterial adaptability [26]. When the NtrX gene is disrupted, *Neisseria gonorrhoeae* cannot efficiently regulate the expression of its respiratory-related genes. This results in the bacteria's ability to breathe under aerobic and anaerobic conditions being affected, which may affect their ability to survive, reproduce and adapt to different environments. The NtrYX system of *Neisseria gonorrhoeae* is speculated to be a key regulator of respiratory enzyme expression, and plays a key role in the adaptability of the bacterium, Such as nitrite and - Nitric oxide reductase and cytochrome c oxidase (CcoP subunit), and control biofilm formation and virulence [26]. However, our results also show that in N. gonorrhoeae, NtrYX plays a broader role in regulating respiratory gene expression, as cytochrome cbb3 and cytochrome c peroxides in Ng-ntrX mutants The expression of enzymes was also down-regulated [26].

Sinorhizobium melilotiis a Gram-negative bacterium that can be symbiotic with leguminous plants such as alfalfa. This bacterium can form a symbiotic relationship with the roots of alfalfa, through the Nitrogen fixation converts nitrogen in the air into ammonia nitrogen that can be used by plants. There are 102 TCS genes (44 HK and 58 RR) in *S. meliloti* 1021 [27]. Among them, there have been seven HK genes and 12RR genes were studied through gene mutation, including TCS NtrB/NtrC. This TCS is required for nitrogen fixation and nif gene expression in *Klebsiella pneumoniae* and free *Azorhizobium caulinodans* [28]. However, NtrB/NtrC is not required for symbiotic nitrogen fixation in *S. meliloti* or *Bradyrhizobium japonicum*, suggesting the presence of unknown regulators involved in symbiotic nitrogen

metabolism in these rhizobia [28]. Through research, we found that the response regulatory protein NtrX of *S. meliloti* is not a histidine kinase NtrY, and the response regulatory protein NtrX acts as a independent of histidine kinase NtrY regulates the production of succinoglycans and the formation of flagella, and is involved in the regulation of exopolysaccharide production, motility and symbiotic relationship with alfalfa. A plasmid insertion mutant of NtrX forms sticky colonies and overproduces succinoglycan, an exopolysaccharide, by upregulating its biosynthetic genes [29]. Research findings indicate that NtrX of *S. meliloti* is a novel regulator of succinoglycan production and motility that is not genetically coupled to NtrY [29]. Our findings not only indicate that NtrX is a new player in the regulation of EPS production and motility, but also prove that the NtrX gene has pleiotropic effects in rhizobia [29]. In *Sinorhizobium meliloti*, deletion of NtrYX leads to increased exopolysaccharide synthesis, reduced tolerance to salt and detergents, and changes in cell morphology [29, 30].

In addition, in *Bradyrhizobium diazoefficiens*, the NtrYX system plays a key role in symbiotic nitrogen fixation in soybean plants and cbb3 oxidase expression in the bacteria [31].

Bacteria sense and respond to environmental changes through a two-component system to maintain their survival and reproduction. Studying the functions of NtrYX two-component systems in different bacteria will help us gain a deeper understanding of how bacteria adapt to various environmental stresses, such as temperature, pH values, nutrient availability, etc. While providing a better understanding of the formation and maintenance mechanisms of symbiotic relationships between bacteria.

1.4.2 Synergy between NtrYX two-component system and other two-component systems

Caulobacter crescentus is a free-living α-proteobacteria that lives in freshwater and soil environments. It encodes both ChvGI and NtrYX systems [32, 33]. ChvGI of *Caulobacter crescentus* can activate and regulate the transcription of RNA ChvR,

thereby post-transcriptionally inhibiting the TonB-dependent receptor gene ChvT [34]. A recent study showed that NtrX is phosphorylated during stationary phase in specific media due to acidification [35]. Deletion of ChvG and ChvI was found to produce significant growth defects in certain media. Taking advantage of this flaw, we discovered ChvGI and NtrY, NtrX. There are significant genetic interactions between and previously uncharacterized genes NtrZ. epistasis analysis provides evidence that unphosphory lated NtrX responds to the lack of ChvG or ChvI cells are harmful. We defined the transcriptional regulator of ChvI and found that it significantly overlaps with genes regulated by NtrX. We conclude that ChvGI and NtrYX interact at multiple transcriptional levels to both synergistically and relatively regulate the growth of *Caulobacter crescentus* in specific media [16].

In *Brucella spp*, the two-component system PrrBA and NtrYX synergistically regulate its adaptability to low-oxygen environments [25]. According to the results, the Brucella abortus ntrY-prrB double mutant strain is defective in in vitro growth and virulence in macrophages and mice, whereas these phenotypes are not significantly affected in the single mutant strain [25]. This suggests that PrrBA and NtrYX have a synergistic effect due to the compensatory function of inducing denitrification and high-affinity cytochrome oxidase genes under microaerophilic conditions [24]. Cytochrome cbb3 expression of oxidase and denitrification pathway genes under microaerophilic conditions in NtrY and PrrB were reduced in the single mutant strain, but surprisingly, they were almost completely turned off in the double mutant [36]. This result suggests that the two-component system PrrBA and NtrYX has a compensatory function in controlling the expression of these genes under low oxygen tension [25]. The status of the nitrogen source in the host cell is crucial for the intracellular survival of *Brucella* [28]28, and the NtrBC system plays an important regulatory role in nitrogen assimilation [37]. Thus, activation of NtrBC by PrrBA and NtrYX may expand the regulatory network used by Brucella to adapt to intracellular life and may explain the defect in growth of the double mutant in an aerobic environment [25]. In addition, we found that the ntrY-prrB double mutant strain was sensitive to oxidative stress [25]. *Brucella* is exposed to reactive oxygen

species produced by the oxidative burst of macrophages and its own aerobic metabolism [38]. Therefore, PrrBA and NtrYX may be involved in regulating genes required to overcome oxidative stress. Further experiments are needed to confirm this hypothesis [25].

1.5 Introduction to Bioinformatics

Bioinformatics is an emerging discipline that acquires, analyzes, processes and applies biological information. It originates from the intersection of biology and computer science, applied mathematics, computer science and physics. It has developed rapidly in recent years. Showing broad application prospects in genetic analysis, tumor diagnosis and treatment, agricultural scientific research and other directions [39-40]. It is one of the major frontier fields of life sciences and natural sciences today, and will also be one of the core fields of biomedicine in the 21st century. Its main research contents are genomics and proteomics [41].

Bioinformatics is widely used in many fields. In genomics, by analyzing genome sequences, people can understand the relationship between genes and diseases, discover new disease genes, and provide scientific basis for disease prevention and treatment. In proteomics, bioinformatics helps people understand the structure and function of proteins, explore the interactions between proteins, and reveal the regulatory mechanisms of proteins. In addition, bioinformatics also plays an important role in the fields of metabolomics, transcriptomics, epigenetics and other fields.

Using bioinformatics to study NtrYX two-component signal transduction system helps us understand how bacteria adapt and respond to environmental changes, and may also provide insights into future biotechnology applications New ideas and methods.

There are many software and tools in bioinformatics that we need to use to help us conduct research. For example, in bioinformatics it is usually necessary to perform sequence alignment on DNA, RNA or protein, and using NCBI blast we can quickly

Alignment,It finds similar known protein structures by comparing the similarity between the known protein sequence and the target protein sequence.

1.6 The purpose and significance of the study

In-depth understanding of the structure, function, regulatory mechanism and possible environmental adaptability of the NtrYX two-component signal transduction system in *Azobacterium* through bioinformatics methods , not only helps to gain a deeper understanding of the structure and function of the system, promotes the research of *Azobacteria*, further reveals its nitrogen fixation mechanism and application potential, but also helps to promote the research and development of *Azobacteria*, two-component systems and bioinformatics, providing strong support for a deep understanding of the biological significance and application value of this system.

This research can also provide new ideas and methods for the field of biotechnology. For example, the NtrYX system can be used to regulate the nitrogen-fixing ability of *Azobacterium* and improve its application effect in agricultural production; this system can also be used to study the biology of other microorganisms properties, providing more possibilities for the field of biotechnology. It is also possible to discover new biological phenomena or functions, thereby exploring their application potential in biotechnology, agriculture, environmental protection and other fields.

CHAPTER 2
OBJECT, PURPOSE, AND METHODS OF THE STUDY

2.1 Sequence search

The sequence of the two-component nucleic acidof *Cereibacter azotoformans*, the subject of this project, has been sequenced by our laboratory. To learn more about the details of these two sequences and to determine whether these two sequences had been obtained and submitted to the database by other researchers, we performed a database search. Use the sequence search tool Basic Local Alignment Search Tool (BLAST) to place the two sequences in the nucleic acid sequence database GenBank (https://www.rcsb.org/) and the protein sequence database UniProt (https://www.uniprot.org/) to perform sequence similarity searches. GenBank is a DNA sequence database established by the National Center for Biotechnology Information (NCBI) in the United States. It obtains sequence data from public resources, mainly provided directly by scientific researchers or derived from large-scale genome sequencing projects. UniProt is the most informative and comprehensive protein database. It is formed by integrating data from three major databases: Swiss-Prot, TrEMBL and PIR-PSD. His data mainly comes from protein sequences obtained after the completion of the genome sequencing project. It contains a large amount of information on the biological functions of proteins from the literature. BLAST is a bioinformatics algorithm used to compare and identify similarities between nucleotide or protein sequences. In GenBank databaseuse BLAST's blastx to perform a protein sequence search on known nucleic acid sequence entries (for nucleotide queries, translate them into proteins and search the protein database), and retrieve proteins similar to the target protein. sequence. In the UniProt database, the protein sequence entries converted from the nucleic acid sequences are searched through BLAST's blastp.

2.2 Structural domain analysis

Structural domains refer to independent structural units with specific functions within a protein molecule. They are often closely related to the functional properties of the protein. Analyzing the structural domains of protein sequences helps us understand the structural characteristics of the protein and its relationship with function. SMART (Simple Modular Architecture Research Tool,http://smart.embl.de/) is a tool for identifying and analyzing protein sequences based on conservation of protein sequences. Online tools for domains and functional sites. The basic steps are as follows: 1) Visit the SMART website; 2) Enter the protein sequence, which can be a single amino acid sequence or multiple sequences, and the sequence format is FASTA; 3) Select database and settings, choose the appropriate database according to your needs,SMARTallows users to adjust some search parameters to optimize results; 4) performs a search. Use SMART to analyze the protein sequences of NtrY and NtrX respectively, and obtain the corresponding protein domain analysis results.

2.3 Construction of 3D structure

2.3.1 3D structure retrieval

Place NtrY and NtrX protein sequences in PDB and AlphaFold protein structure database respectively. Search to determine whether the corresponding protein three-dimensional structure has been experimentally determined or theoretically predicted. Protein Data Bank (PDB) database (https://www.rcsb.org/)is the most famous and complete database in the world created in 1971 The three-dimensional structure database of biological macromolecules mainly contains structural data determined by experimental methods such as X-ray crystallography, nuclear magnetic resonance spectroscopy, and cryo-electron microscopy. AlphaFold Protein Structure Database (AlphaFold DB, https://alphafold.ebi.ac.uk) Biological macromolecules developed by DeepMind Company based on artificial intelligence technology, a high-precision three-dimensional structure prediction tool whose data reliability has been fully verified. If the database search can obtain a result that is

100% consistent with the target sequence, it means that the three-dimensional structure of the target protein has been experimentally analyzed or theoretically predicted, and the corresponding structure file can be directly downloaded for sustainable research.

2.3.2 Homology modeling

From the UniProt database, there are currently approximately 560,000 amino acid sequence information, while there are only 190,000 structural information stored in the PDB database. Such a gap shows the fact that even if we already have very powerful structural analysis tools, the speed of obtaining structures by experimental methods alone cannot keep up with the speed of sequence increase. Therefore, it is very necessary to rely on new technical means to complete the transformation from sequence to structure. Based on different prediction principles, protein structure prediction methods are also divided into several types, mainly including traditional homology modeling, fold recognition, methods such as ab initio predictionand deep learning algorithms based on machine learning. Homology modeling, also called comparative modeling, is currently the most reliable and very practical method for predicting protein structure. The premise of homology modeling is that the structure of one or more homologous proteins is known. When the sequence homology between two proteins is higher than 35%, their three-dimensional structures are generally considered to be basically the same; for proteins with sequence homology below 30%, it is difficult to obtain an ideal structural model. Fold identification is based on the prediction that proteins with very low sequence homology may have structurally identical foldons. Ab initio prediction means predicting the tertiary structure directly from the sequence itself, which is suitable for situations where there is neither a homologous protein with a known structure nor a foldon with a known structure. The basic idea is to combine known methods with computational chemistry and traditional physics methods, using simplified protein models and average potential fields derived from proteins with known structures to theoretically calculate the structure of the protein. Based on the relatively ideal homologous protein

structure templates found in the database for the proteins of this system, in this study we used the homology modeling method to construct the protein structures of NtrY and NtrX. The modeling platform uses SWISS-MODEL modeling server.

2.3.3 Lagrange diagram to evaluate modeling results

Models obtained by homology modeling usually require quality analysis of the model to ensure its reliability. Here we use the PROCHECK tool in SAVES v6.0 (https://saves.mbi.ucla.edu/) for analysis. PROCHECK evaluates whether the protein structure is reasonable by drawing a Ramachandran Plot to display the distribution of each amino acid. It displays the conformation of the protein chain based on the two angles of φ (phi) and ψ (psi) of the amino acid residues.

In the Rascher diagram, according to the distribution of φ and ψ angles of amino acid residues, it can be divided into four regions: the most favorable region (favorable region), the allowed region (allowed region), the generally allowed region (generally allowed region) and Disallowed region. A high-quality protein model should have most of its amino acid residues located in the most favorable regions, while avoiding too many amino acid residues in disallowed regions.

2.4 Molecular docking

Molecular docking is an important step in the study of biomolecular interactions. Molecular docking is a computational simulation method used to predict interactions between small molecules (such as ATP) and large molecules (such as proteins). It can help us predict the binding mode, affinity and possible biological activity between small molecules and biological macromolecules. CB-Dock2 (https://cadd.labshare.cn/) as a molecular docking tool, it provides us with an efficient and accurate method to simulate and Study these interactions and provide detailed binding mode and affinity predictions. In this topic, we use CB-Dock2 as a molecular docking tool, simulate ATP and NtrY proteins to explore the binding mode and affinity between them.

2.5 Protein-protein interactions

The interaction between proteinsnot only determines the specific functions of proteins in cells, but also involves complex signaling networks, thereby affecting the overall physiological and pathological processes of organisms. In order to study the interaction between NtrY and NtrX proteins, we used the ClusPro (https://cluspro.bu.edu/) toolto predicttheinteraction between NtrY and NtrX proteins. ClusPro is a tool for predicting the three-dimensional structure of protein-protein complexes. We upload the PDB files of NtrY and NtrX to ClusPro, and ClusPro will automatically start to simulate the mode of action of NtrY and NtrX proteins. ClusPro will generatethe complex structure of NtrY and NtrX proteins based on factors such as charge distribution and possible interaction sites.

CHAPTER 3
EXPERIMENTAL PART

Sequence search found that the sequences of NtrY and NtrX of this system have sequence entries with 99% and 100% consistency respectively in the GenBank database. The retrieval numbers are WP_238823272 and WP_011908515, respectively representing *Cereibacter azotoformans* is a PAS domain-containing sensor histidine kinase and sigma-54-dependent transcriptional regulator. UniProt database searches found sequence entries with 91.1 and 94.7% identity respectively: Q3J2L3 and Q3J2L4. The corresponding entry protein of Q3J2L3 has the function of catalyzing histidine phosphorylation. The molecular functions of the protein corresponding to Q3J2L4 include activator and binding to DNA; biological processes include transcription, transcriptional regulation and two-component regulatory system; ligands include binding to ATP and nucleotides. This information provides a wealth of information for us to understand the NtrYX two-component system of this system.

3.1 Domain analysis of proteins

Through database search and sequence analysis, we finally determined the protein sequences of NtrY and NtrX, as follows:

>NtrY

MRTAARSNTWIKLARWRRQRRFQNAAAFGLVMLGPVLALATFLALGPM
NQGANSPGLRFVLLADLVYVLVVAALVIARIARMVSDRRSQSAGSRLHLRLV
GTFAVLALVPTVLVAVFAMLTVNVGIEGWFSDRVRQVVGASLNAAEAYQEE
HRRDLIEDAEALAAYLNVAKQSTFFLRDDQLRPLLSQGQDRIQRGLREAFLID
GGGELRTRGERSYLFDFEKPAPEDIERARSGETALIQDWANNEFRALVHLEAF
PDRFLYVSRTVDGSILSLLDDTRETVVLYHQLEAERGRMLFEFGLLYLGFALI
LILAAVWLGLWFAERLSRPVGRLAGAAQRVGAGDLDVQVVEEEGDDEIAM
LGRLFNQMTRQLKGQRDALMDNNRQTERRRRLFDSVLSSVTAGVIGLDAVG
QVDFINRAAERLLELPEAGNLPLSVAVPEFAALFERLRETGAAVQEEIRLIRK

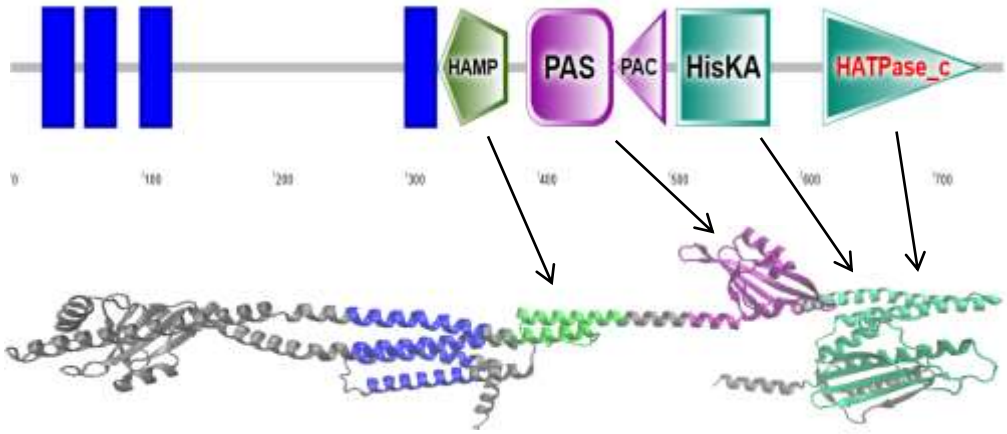
GRMESLLVRMSPRRTESGRLEGYVVAFDDVTDLVSAQRMAAWGDVARRIA
HEIKNPLTPIQLSAERIKRKFRPLVGEQAGDLDQYADVIIRQTNDLRRIVDEFS
KFARMPEPDRREADLVKLVRDAVLLQEAGQPGVRIRAALPAEAWLIDIDTT
MIGQALTNLMKNAGEAIEARAENEQADWRGEIRVSLAVDEDQALIRISDNGT
GLPPDRTRLFEPYVTTREKGTGLGLPIVKKIIEEHGGILTLADAEAIFEDGHRG
AMAEIRLPRILRSRARAARASEAILEEK

>NtrX

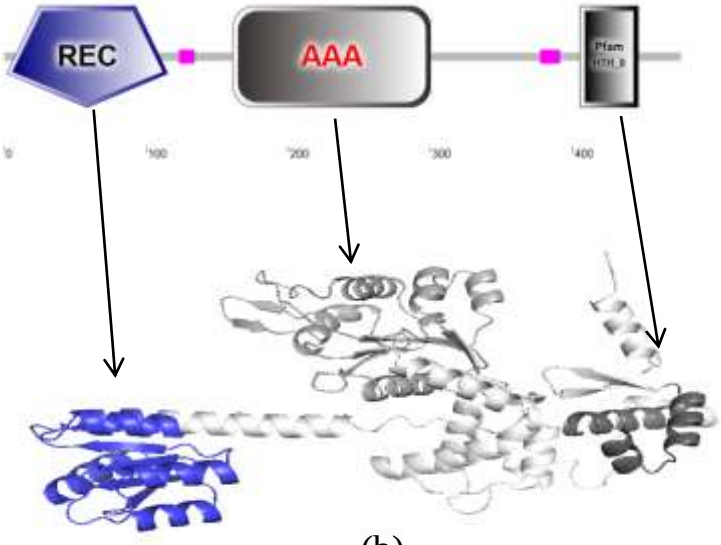
MSSILIVDDERDIRELVGDILRDEGFQIRLAANSDECMAAINAEPPALMILDI
WLKDSRMDGIDILKRTKRDNPDVPVVIISGHGNIEIAVAAIKQGAYDFIEKPFN
IDQLMVVVQRAMETARLRRENSELRRRDTSTAEMLGASPAFRLLRSQLEKVT
KSNGRVMLSGPAGSGKEMAARFIHSNSGRAGGPFVSVSSATVQPDRMEEVL
FGRETPDRGIEQGLLEQAHGGIVYFDEVADMPPGTQSKILRVLTEQQFTRQG
GTDKVRVDLRVISSTTRDLRAEIAAGRFRQELYDRLNVVPIEVPALADRREDI
PMLARHFIEMFHRSQGLPLRSLTGEAEAMLQTMLWPGNVRQLRNVIERVLIL
GDGSGPIEARELPGNDAPGEEGRLILGGALATLPLREARELFEREYLLTQINRF
GGNISRTAAFVGMERSALHRKLKSLGVVTTAKGGSRLARIEDDFEDEEEALG
APD

The two sequences were analyzed using SMART for domain analysis. The analysis results are shown in Figure 3.1, where a represents NtrY and b represents NtrX. As shown in Figure 3.1a, the NtrY protein contains the transmembrane region, PAS, PAC, HAMP, HisKA and HATPase_c6 domains. Thetransmembrane region is composed of fourα-helical regions, which is part of the two-component signal transduction pathway. Research shows thattransmembrane regionpasses Structural changes in periplasmic binding proteins are transmitted to kinases and methylated receptor domains in the cytoplasm to regulate receptor phosphorylation or methylation. The HAMP domain may regulate receptor phosphorylation or methylation by transmitting structural changes in the periplasmic binding protein to kinases and methylating the receptor domain in the cytoplasm. PASdomains occur in archaea, bacteria, and eukaryotes and are involved in many signal transduction proteins, which are often used as signal sensors. PAS domains are often associated

with PAC domains. They appear to be directly linked, and together they form the conserved 3D PAS fold. PAC domains occur at the C-terminus of a subset of all known PAS domains, which contributes to the folding of the PAS domain. It is worth mentioning that some studies have shown that the 3D PAS domain is not necessary for some two-component systems, such as CpxA, a two-component system that serves as an environmental pressure sensor in E. coli, there is no 3D PAS domain in the structure [42]. Therefore, the function and role of the 3D PAS domain in signal transduction in the two-component system deserves further study.



(a)



(b)

Figure 3.1 – Domain analysis of NtrYX: a. NtrY, b. NtrX

HisKAis the dimerization and phosphorylation receptor domain of histidine kinase. Histidine kinases typically have a N-terminal ligand binding domain and a C-

terminal kinase domain, but may also Other domains exist. The kinase domain is responsible for autophosphorylation of histidine with ATP, phosphorylation of aspartate from the kinase to the response regulator, and phosphoaspartyl The phosphate transfer back to ADP or the phosphate transfer of water. The homodimer domain includes sites for histidine autophosphorylation and phosphate transfer reactions. The structure of the homodimer domain consists of a closed four-helix bundle with a left-handed twist composed of two identical α-hairpin subunits. The HATPase_c domain is present in a variety of ATP-binding proteins, and the fold of this domain consists of two layers, α/β, containing an 8-strand mixed β sheet.

Figure 3.1b shows that the NtrX protein domain consists of REC, AAA and PfamHTH_8 REC is a CheY-homologous receptor domain that contains a phosphoacceptor site that is phosphorylated by a histidine kinase homologue and receives signals from sensors in a two-component system. AAA are ATP enzymes involved in various cellular activities. Generally, AAA+ domain can be divided into two structural subdomains, one N-terminal Ploop NTP enzyme α-β-α subdomain, connected to the smaller C terminal full α sub domain. bacterial regulatory protein (FIS)is referred to as HTH_8, is a Pfam domain. It activates ribosomal RNA transcription and is involved in the upstream activation of the rRNA promoter. At its Cterminus, FIS encodes a helix-turn-helix (HTH) DNA binding motif. The sequence shows high similarity to the HTH motif of other more primitive bacterial transcriptional regulators. Some of its functions include inhibition of DNA replication starting from the OriC site, and promotion of Hin-mediated inversion of DNA.

3.2 3D structure analysis

When constructing the 3D structures of NtrY and NtrX proteins, first search the PDB database (https://www.rcsb.org/) to find whether there is already a protein with NtrY and NtrX. Proteins with similar sequences or similar structures to the NtrX protein were analyzed and stored in the database. This similarity search is based on amino acid sequence similarity, structural domain homology, and possible biological functions. Download selected structural data from PDB database using protein

structure visualization software (such as PyMOL, Chimera, etc.) Open the downloaded coordinate file to view and analyze the 3D structure of the protein. Use NtrY and NtrX protein sequences to search for templates with higher similarity to them in PDB. The search found that the PDB access number of the template with the highest similarity to the NtrY protein is 3A0R (identity 29%), and the PDB access number of the template with the highest similarity to the NtrX protein is 5M7N (Consistency 56%). Among them, in 3A0R this template contains the structure of another protein, which may be useful for studying protein-protein interactions. helpful. In the AlphaFold protein structure database, we retrieved two structural data with 100% consistency with the two target sequences. The search numbers are AF-A4WSL9-F1 and AF-A4WSM0-F1 respectively, indicating the structures of these two proteins. Although the experimental method has not been analyzed yet, it has been predicted by AlphaFold. The corresponding experiments and predicted structures are shown in Figures 3.2.

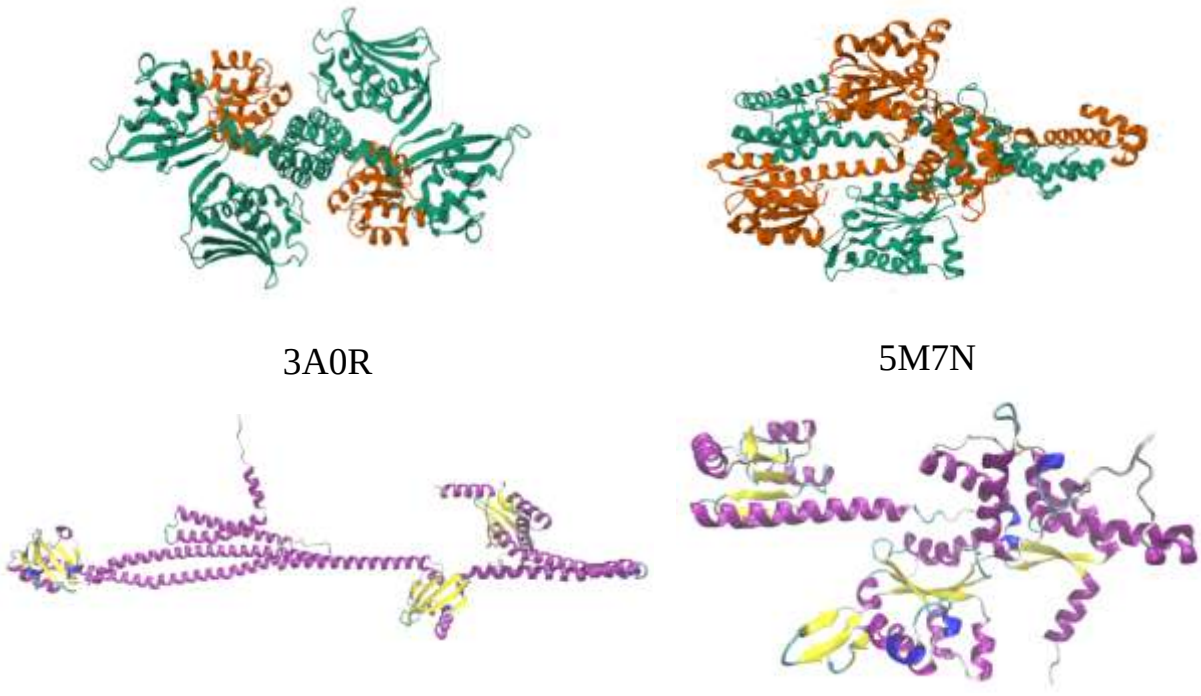


Figure 3.2 – The 3D structure of proteins related to NtrY and NtrX searching from databases

3.2.1 Homology modeling

Although the protein structure of the AlphaFold protein structure database entry AF-A4WSL9-F1 is 100% consistent with NtrY, its structure is a monomer, and experiments have proven that the intracellular part of NtrY exists in the form of a dimer in vivo. Normal biological functions can only be exerted in the dimer state. Therefore, in order to ensure the accuracy of the next experimental data, we must construct the dimer structure of NtrY. We respectively tried to directly construct dimers through Z-dock (https://zdock.wenglab.org/) using the AF-A4WSL9-F1 structure as a template, and constructed NtrY dimers through homology modeling through SWISS-MODEL. aggregate. Through comparison and analysis of the modeling results, we finally determined that the structure of the NtrY dimer constructed by homology modeling is more reasonable (Fig. 3.3). In the subsequent process, we used this Conduct experimental research based on the structure.

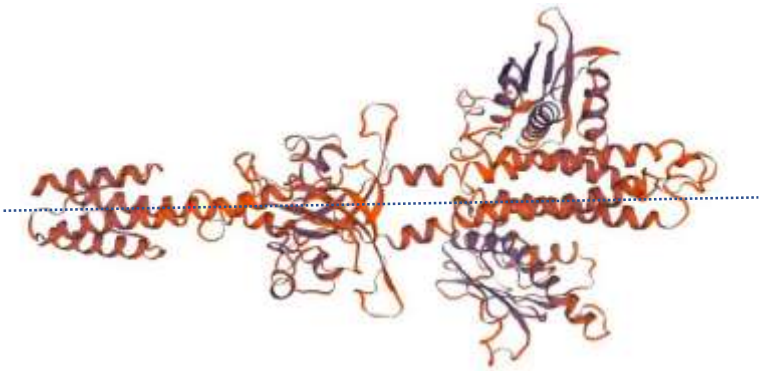


Figure 3.3 – Constructing NtrY intracellular dimers through homologous modeling

3.2.2 Lagrange diagram to evaluate modeling results

The model obtained by homology modeling may have some biases or errors. We usually need to perform quality analysis on the model to ensure its reliability. Here we use SAVES v6.0 (https://saves.mbi.ucla.edu/) The PROCHECK tool is used for analysis. PROCHECK evaluates whether the protein structure is reasonable by drawing a Ramachandran Plot to display the distribution of each amino acid. It

displays the conformation of the protein chain based on the two angles of φ (phi) and ψ (psi) of the amino acid residues.

In the Rascher diagram, according to the distribution of φ and ψ angles of amino acid residues, it can be divided into four regions: the most favorable region (favorable region), the allowed region (allowed region), the generally allowed region (generally allowed region) and Disallowed region. A high-quality protein model should have most of its amino acid residues located in the most favorable regions, while avoiding too many amino acid residues in disallowed regions.

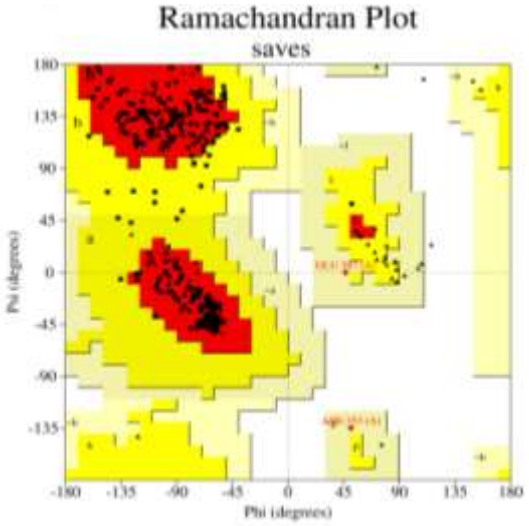


Figure 3.4 – Ramachandram plot analysis of NtrY intracellular dimer

As can be seen from the Lagrange diagram (Fig. 3.4), the template distributes more than 90% (93.6%) of the residues in the most favorable region, and in Residues in disallowed regions are all 0. This indicates that most of the amino acid residues of the protein model are in energetically stable and biologically common conformations. This shows that the protein model constructed is of high quality and reasonable, and the structural characteristics of the model are consistent with the statistical data in the known protein structure database.

3.3 Protein-ligand interaction analysis

3.3.1 Analysis of NtrY extracellular signal sensing region

Molecular docking is an important step in drug design and the study of biomolecular interactions. Molecular docking is a computational simulation method used to predict interactions between small molecules (such as ATP) and large molecules (such as proteins). It can help us predict the binding mode, affinity and possible biological activity between small molecules and biological macromolecules. CB-Dock2 (https://cadd.labshare.cn/), as a molecular docking tool [43-44], provides us with an efficient, Accurate methods to model and study these interactions and provide detailed binding mode and affinity predictions.

The extracellular part of NtrY is the signal sensing region of the two-component signal transduction system, but so far, its signaling molecules have not been experimentally identified. Here we utilize bioinformatics tools to predict potential signaling molecules. Combine domain and the three-dimensional structure of NtrY, we can locate the sequence range of the NtrY extracellular signal sensing region from 122 to 299, extracted from the AF-A4WSL9-F1 structure 122 to 299 sequence interval structure, and then submit the structure to the COACH server. COACH is a metaserver method for protein ligand binding site prediction. Starting from a given target protein structure, COACH will generate complementary ligand binding site predictions using two comparison methods, TM-site and S-site, which combine specific substructure and sequence map comparisons. Identify ligand binding templates in the BioLiP protein functional database [45-46]. The predicted results are shown in Figure 3.5. As shown in the figure, the active region of the NtrY extracellular signal sensing region is located in the flexible region of the domain composed of 2 helices and 1 fold. The signal molecule is compared to the citrate ion (Citrate anion, molecular formula: $C_6H_5O_7$), which is related to the activity Pockets of randomly coiled residues interact. By analyzing the interaction between the two using LigPlot+ software, it can be found that Arg215 and Arg198 residues have strong hydrogen bond interactions with small molecules, forming 3 and 2 hydrogen bonds respectively, with hydrogen bond lengths ranging from 2.95 to 3.11. between Å. In

addition, Tyr217 and Asp220 have obvious hydrophobic interactions with small molecules.

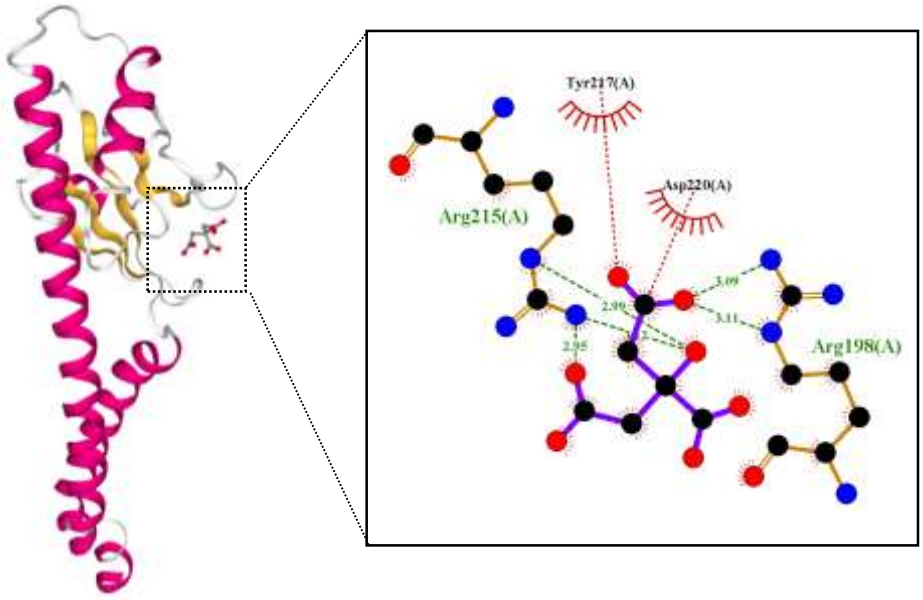


Figure 3.5 – Structure of NtrY extracellular signal sensing region and protein substrate interactions predicted by COACH

3.3.2 Analysis of interaction between NtrY and ATP

In the process of two-component signal transduction, the phosphorylation of histidine kinase plays a relay role. On the one hand, it receives upstream signal stimulation, and on the other hand, it activates downstream signaling pathways by changing its structure, thereby regulating gene expression. The phosphorylation of histidine kinase starts from the interaction between NtrY protein and ATP. Therefore, the study of the interaction between NtrY protein and ATP is important for understanding the phosphorylation of histidine kinase in the NtrYX two-component system. significance. Molecular docking of NtrY dimer and ATP through CB-Dock2, the docking results analyze the interaction between protein and ligand through LigPlot+ software, and the resultsare shown in Figure 3.6.

After the dimer of NtrY protein binds to ATP molecules, it forms two active pockets (region A and region B), both of which are located in the HisKA domain of

NtrY protein. This also shows that the HisKA domain is the autophosphorylation region of NtrY protein and ATP, which is consistent with the experimental research

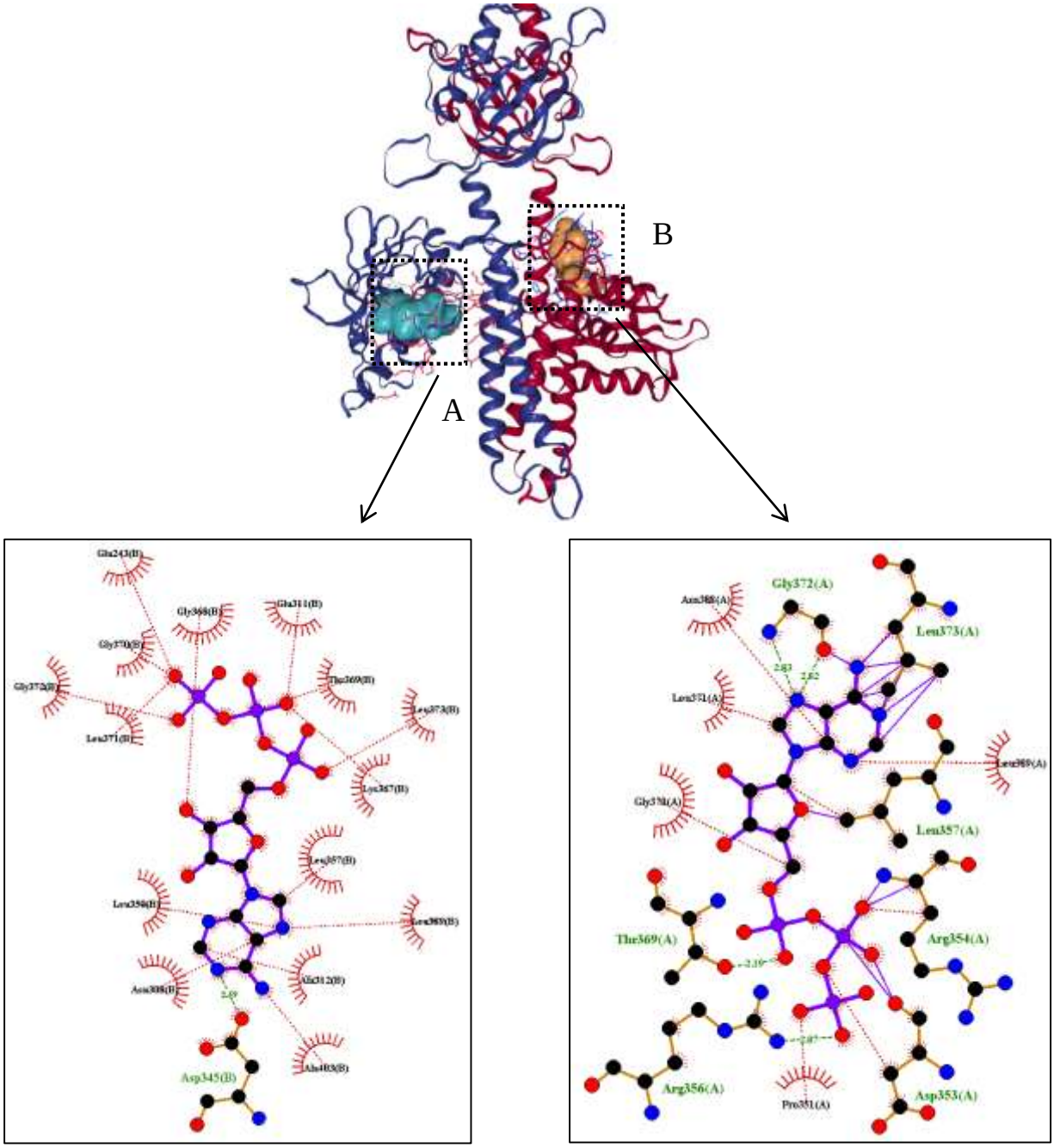


Figure 3.6 – Interactions of NtrY with ATP

results [42]42. However, it is worth noting that the distances between regions A and B from theαhelix of the HisKA domain are not the same, and region B is significantly closer. LigPlot+ interaction force analysis found that the modes of action of the two

active regions are also different. The interaction between ATP and NtrY is significantly stronger inB region Yu is in area A. In the B region, there are multiple residues (Gly372, Thr369, Arg356, etc.) that form strong hydrogen bonds with ATP, while in the A region, only one residue, Asp345, participates in the formation of hydrogen bonds. This result shows that the structure of the NtrY dimer is deformed, and the deformation of the structure is caused by external stimulation signals. This has been confirmed by experiments. The crystal structures of the activated and inactive states of the two-component CpxA have been analyzed [42]42. In the activated state, it can be clearly seen the configuration of the dimer changes, and the HisKA region of one of the dimers is closer to the histidine residue on the α-helix, which puts the transfer of the phosphate group of ATP within an effective distance range.

3.3.3 NtrX interacts with substrate

NtrX is the response regulator of the NtrYX two-component system. After it is phosphorylated, it undergoes a conformational change and becomes an activated state. NtrX in the activated state can combine with bacterial genes, and then Regulate the expression of bacterial-related genes to respond to external signal stimulation. After response regulation is completed, NtrX is dephosphorylated and returns to its inactive state. This process involves the interaction between NtrX and the substrate small molecule ADP. Understanding the interaction mode between the two is important for understanding the phosphorylation and dephosphorylation mechanisms of NtrX is very important meaning. Submit NtrX three-dimensional structure files to the COACH server, and the prediction results are shown in Figure 3.7. Residues Ser173, Gly174, Val354 and Arg355 play an important role in the docking process of NtrX and ADP, generating strong hydrogen bonds with ADP. The length distribution is between 2.54-3.32 Å. Among them, residue Ser173 forms a total of three hydrogen bonds with the phosphorus atoms and nitrogen atoms at both ends of ADP, which plays a particularly important role in the stability of the ADP structure.

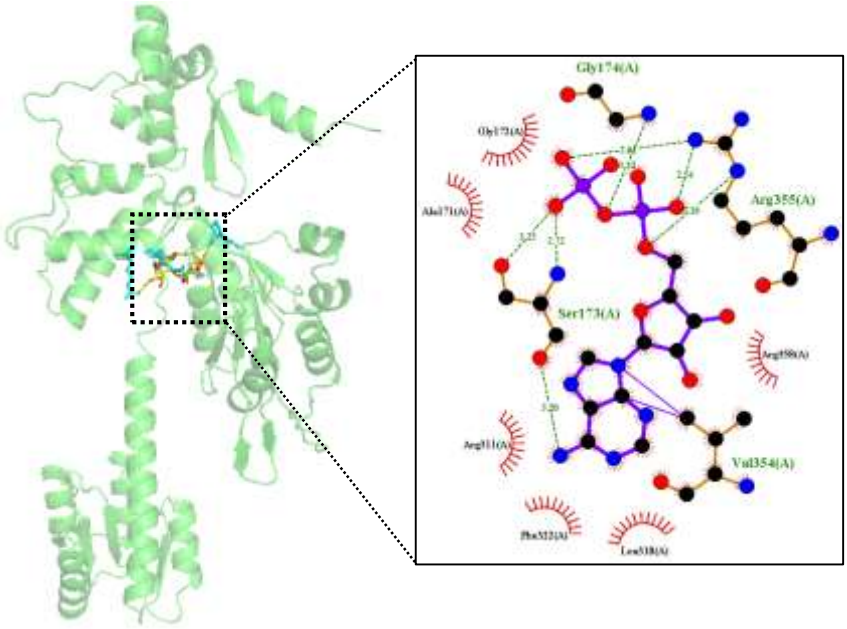


Figure 3.7 – Interactions of NtrX with ADP

3.3.4 α-helical interaction of dimer HisKA domain

Koh et al. found that the activation of the HisKA domain of the BceS two-component system is triggered by helical rotation in its body [47]47, between the α-helices of the dimer The interaction plays a key role in its rotation. To this end, we analyzed the interaction between the main α-helices of the HisKA domain in the NtrYX two-component system. Extract the α-helix under the active HisKA domain from the homology modeling dimer structure, and then use LigPlot+ to analyze the interaction between the two stripsα-helix force. In order to compare the changes in the interaction force between the two α-helices in the inactive and active states, we used the PdbViewer software to use the gromos96 43B1 force field to minimize the energy of the α-helix in the active state, so that the two the structure of the α-helices tends to be in an inactive state, and then LigPlot+ was used to analyze the interaction force of the two optimized α-helices. The results are shown in Figures 3-8. From this we can see that in addition to hydrophobic interactions, there are also strong hydrogen bonds between some residues between the twoα-helices. One hydrogen

bond is formed between Lys195(A)-Ser245(B) and Lys209(A)-Asp245(B), and two hydrogen bonds are formed between Arg231(A)-Asp244(B) and Arg238(A)-

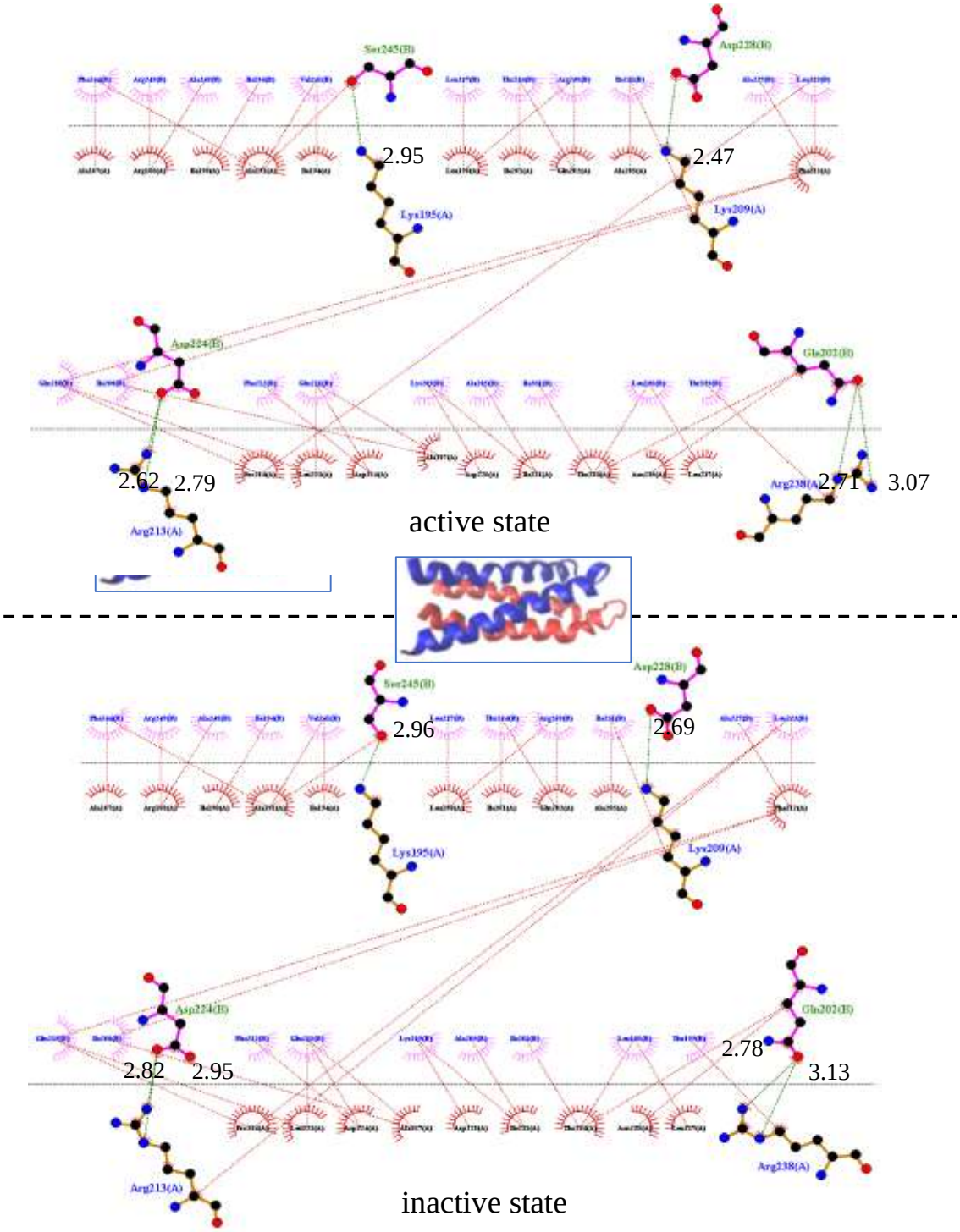


Figure 3.8 – Interactions of active/inactive dimer HisKA Domain α

Gln202(B). a hydrogen bond. In the active state, the length of these hydrogen bonds is between 2.47-3.07Å, and in the inactive state, the lengths of these hydrogen bonds are relatively longer, distributed between 2.69-3.13Å. This shows that under the

stimulation of external signals, the interaction between the main α-helices of the HisKA domain is strengthened, which in turn causes the structure of HisKA to change from an inactive state to an active state, providing a basis for the phosphorylation of histidine kinase. spatial distance. There are currently no experimental reports on this finding. In addition, it can be seen from the figure that residues Ser245(B), Arg238(A) and Gln202(B) have undergone obvious structural space changes before and after activation.

3.4 Protein-protein interactions

Studying the interactions between proteins is the key to understanding their functions and signal transduction. These interactions not only determine the specific functions of proteins in cells, but also involve complex signaling networks, thus effects the overall physiological and pathological processes of organisms. The interaction between NtrY and NtrX proteins is a key step in the phosphorylation of response regulator (RR, here NtrX). After NtrX is phosphorylated, it undergoes a conformational change and becomes an activated state. In the activated state, NtrX can combine with bacterial genes and then regulate the expression of bacterial-related genes to respond to external signal stimulation. Therefore, studying the interaction between NtrY and NtrX proteins is of great significance to our understanding of the phosphorylation mechanism of NtrX.

Z-dock is a tool for predicting the three-dimensional structure of protein-protein complexes. Upload the PDB files ofNtrY and NtrX to Z-dock. The prediction results are as shown in the figure as shown in Figure 3.9. The results show that there is a tight binding relationship between the PAS regionof the NtrY protein and the REC region of the NtrX protein, which is the NtrYX double group an indispensable link in the signal transduction chain of subsystems. PAS plays an important role insignal recognition and transduction of the NtrYX two-component system of *C. azotoformans*. PAS acts as a signal sensor, recognizing signal changes in the external environment and rapidly transmits the signal to the REC region of the NtrX protein. This transmission process is not only efficient but also accurate, ensuring

rapid diffusion and response of signals within cells. The high sensitivity and specificity of PAS enables it to distinguish between different signaling molecules and accurately identify signals associated with the NtrYX two-component system. This recognition ability is crucial for cells, because it ensures that bacteria respond differently to different signals, thereby regulating the physiological processes of bacteria, allowing bacteria to survive and adapt better Respond to the environment. Responding to signals closely related to one's own survival and development can also avoid unnecessary energy consumption and waste of resources. REC is key to the ability of the NtrX protein to respond to signals. It receives signals from the PAS area and is quickly activated and triggers a series of cellular responses. These reactions play an important role in bacterial growth, development and metabolism. This shows that the PAS and REC regions are the key to signal transduction in the Azobacterium NtrYX two-component system and play an important role in the process of combining NtrY and NtrX proteins. Their tight combination and synergy ensure the rapid transmission and accurate response of signals within cells, providing a solid guarantee for the survival and development of bacteria. This discovery not only deepens our understanding of the two-component signaling mechanism in bacteria, but also provides new ideas and methods for future biomedical research and drug development.

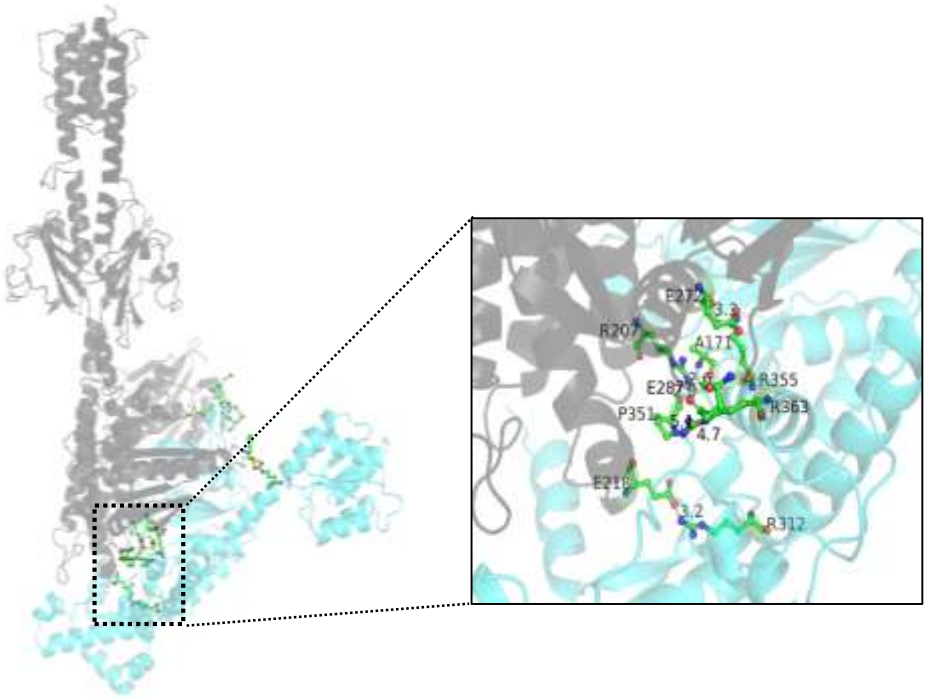


Figure 3.9 – Interactions of NtrY and NtrX proteins

CONCLUSIONS

This study used bioinformatics methods to conduct an in-depth analysis of the NtrYX two-component signal transduction system in *C. azotoformans* and revealed the important role of the NtrYX two-component system in the regulation of nitrogenase activity.

By predicting potential signaling molecules in the extracellular signal sensing region of NtrY, we found that the active region of the NtrY extracellular signal sensing region is located in the flexible region and interacts with the randomly coiled residues of the active pocket. Analyzing their interaction, it was found that residues Arg215 and Arg198 have strong hydrogen bond interactions with small molecules, which indicates that residues Arg215 and Arg198 play an important role in it.

By docking the dimer of NtrY protein with ATP molecules, two active pockets (region A and region B) are formed in the HisKA domain of NtrY protein, which verifies that the HisKA domain autophosphorylates with ATP within the NtrY protein. key areas of action. In area B, the interaction between ATP and NtrY is obviously stronger than that in area A. This indicates that the structure of the NtrY dimer is deformed due to external signal stimulation. Studying the interaction between the α-helices of the HisKA domain, it can be seen that under the stimulation ofexternal signals, the interaction of the α-helices of the HisKA domain becomes stronger, and the HisKA domain changes from inactive to active state. This may have an impact on signal conduction. Analysis of the interaction between NtrX and the substrate small molecule ADP found that residues Ser173, Gly174, Val354 and Arg355 play an important role in the phosphorylation and dephosphorylation process of NtrX protein.

The interaction between NtrY and NtrX proteins is critical to the signal transduction chain. PAS regionrecognizes changes in signals in the external environment and transmits them to REC region. This process is the key to the

combination of NtrY and NtrX proteinsto ensure. It ensures rapid transmission and accurate response of signals within cells.

In addition, simulating the combination of NtrY protein and ATP molecules provides a new perspective for understanding the function of the NtrYX system in bacterial physiological processes, and provides theoretical guidance for experimentally improving the response sensitivity of the two-component system.

Although bioinformatic analyzes provide important information about the structure and function of the NtrYX system, these findings need to be verified by experimental biology methods. This study reveals the components and basic functions of the NtrYX system, but signal transduction in bacteria often involves multiple interacting two-component systems. Future studies can explore the interactions and synergistic effects between the NtrYX system and other two-component systems, and how they integrate multiple environmental signals to regulate bacterial behavior. The role of the NtrYX system in bacterial adaptation to environmental changes also deserves further study. In terms of drug development, key components of the NtrYX system, such as the HisKA domain and REC region, may become targets for new antibacterial drugs. Future research can focus on developing small molecule compounds that can specifically interfere with the function of these regions to control bacterial signal transduction processes.

LIST OF REFERENCES

1. Suwa, Y., Sumino, T., Noto, K.,& et al. (2011). Cereibacter azotoformans gen. nov., sp. nov., a nitrogen-fixing bacterium isolated from soil. *International Journal of Systematic and Evolutionary Microbiology*, 61(7), 1651-1655.
2. Suwa, Y., Sumino, T., Noto, K.,& et al. (2012). Characterization of a nitrogen-fixing bacterium isolated from soil and its potential for bioremediation. *Journal of Bioscience and Bioengineering*, 113(5), 542-546.
3. Liao, H., Li, Y., Wang, H.,& et al. (2020). Nitrogen-fixing bacteria: Diversity, mechanism and application. *Critical Reviews in Biotechnology*, 40(8), 1105-1120.
4. Ormeño-Orrillo, E., & Martínez-Romero, E. (2019). A genomotaxonomy view of the Bradyrhizobium genus. *Frontiers in Microbiology*, 10, 1334.
5. Suwa, Y., Sumino, T., Noto, K.,& et al. (2013). Denitrification activity of Cereibacter azotoformans under oxygen-limited conditions. *Microbes and Environments*, 28(4), 491-494.
6. Ishii, S., Ikeda, S., Minamisawa, K.,& et al. (2011). Nitrogen cycling in rice paddy environments: Past achievements and future challenges. *Microbes and Environments*, 26(4), 282-292.
7. Suwa, Y., Sumino, T., Noto, K.,& et al. (2014). Degradation of aromatic compounds by Cereibacter azotoformans. *Journal of Bioscience and Bioengineering*, 117(3), 310-313.
8. Fuchs, G., Boll, M., & Heider, J. (2011). Microbial degradation of aromatic compounds - from one strategy to four. *Nature Reviews Microbiology*, 9(11), 803-816.
9. Zschiedrich, C. P., Keiddel, V., & Szurmant, H. (2016). Molecular mechanisms of two-component signal transduction. *Journal of Molecular Biology*, 428(19), 3752-3775.
10. Gao, R., & Stock, A. M. (2009). Biological insights from structures of two-component proteins. *Annual Review of Microbiology*, 63, 133-154.

11. Olaya-Abril, A., Luque-Almagro, V. M., Hidalgo-Carrillo, J., Chicano-Gálvez, E., Urbano, F. J., Moreno-Vivián, C., Richardson, D. J., & Roldán, M. D. (2022). The NtrYX Two-Component System of Paracoccus denitrificans Is Required for the Maintenance of Cellular Iron Homeostasis and for a Complete Denitrification under Iron-Limited Conditions. *Int J Mol Sci*, 23(16), 9172.
12. Lemmer, K. C., Alberge, F., Myers, K. S., Dohnalkova, A. C., Schaub, R. E., Lenz, J. D., Imam, S., Dillard, J. P., Noguera, D. R., & Donohue, T. J. (2020). The NtrYX Two-Component System Regulates the Bacterial Cell Envelope. *mBio*, 11(3), e00957-20.
13. Liu, Siyu, MA, Yushu, & WEI, Toshiba. (2023). Function of a potential two-component system protein in oxygluconobacter. *Journal of Microbiology*, 63(8), 3187-3202.
14. Buschiazzo, A., & Trajtenberg, F. (2019). Two-Component Sensing and Regulation: How Do Histidine Kinases Talk with Response Regulators at the Molecular Level? *Annu Rev Microbiol*, 73, 507-528.
15. Stock, A. M., Robinson, V. L., & Goudreau, P. N. (2000). Two-component signal transduction. *Annu Rev Biochem*, 69, 183-215.
16. Stein, B. J., Fiebig, A., & Crosson, S. (2021). The ChvG-ChvI and NtrY-NtrX Two-Component Systems Coordinately Regulate Growth of Caulobacter crescentus. *J Bacteriol*, 203(17), e0019921.
17. Bhate, M. P., Molnar, K. S., Goulian, M., & DeGrado, W. F. (2015). Signal transduction in histidine kinases: insights from new structures. *Structure*, 23(6), 981-994.
18. Marina, A., Waldburger, C. D., & Hendrickson, W. A. (2005). Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *EMBO J*, 24(24), 4247-4259.
19. Tomomori, C., Tanaka, T., Dutta, R., Park, H., Saha, S. K., Zhu, Y., Ishima, R., Liu, D., Tong, K. I., Kurokawa, H., Qian, H., Inouye, M., & Ikura, M. (1999).

Solution structure of the homodimeric core domain of Escherichia coli histidine kinase EnvZ. *Nat Struct Biol*, 6(8), 729-734.

20. Dutta, R., & Inouye, M. (2000). GHKL, an emergent ATPase/kinase superfamily. *Trends Biochem Sci*, 25(1), 24-28.
21. Galperin, M. Y. (2010). Diversity of structure and function of response regulator output domains. *Current Opinion in Microbiology*, 13(2), 150-159.
22. Carrica, M. Del C., Fernandez, I., Martí, M. A., Paris, G., & Goldbaum, F. A. (2012). The NtrY/X two-component system of Brucella spp. acts as a redox sensor and regulates the expression of nitrogen respiration enzymes. *Molecular Microbiology*, 85(1), 39-50.
23. Pawlowski, K., Klosse, U., & de Bruijn, F. J. (1991). Characterization of a novel Azorhizobium caulinodans ORS571 two-component regulatory system, NtrY/NtrX, involved in nitrogen fixation and metabolism. *Molecular General Genetics*, 231(1), 124-138.
24. Roop, R. M. 2nd, & Caswell, C. C. (2012). Redox-responsive regulation of denitrification genes in Brucella. *Molecular Microbiology*, 85(1), 5-7.
25. Carrica, M. C., Fernández, I., Sieira, R., Paris, G., & Goldbaum, F. A. (2013). The two-component systems PrrBA and NtrYX co-ordinately regulate the adaptation of Brucella abortus to an oxygen-limited environment. *Molecular Microbiology*, 88(2), 222-233.
26. Atack, J. M., Srikhanta, Y. N., Djoko, K. Y., Welch, J. P., Hasri, N. H., Steichen, C. T., Vanden Hoven, R. N., Grimmond, S. M., Othman, D. S., Kappler, U., Apicella, M. A., Jennings, M. P., Edwards, J. L., & McEwan, A. G. (2013). Characterization of an ntrX mutant of Neisseria gonorrhoeae reveals a response regulator that controls expression of respiratory enzymes in oxidase-positive proteobacteria. *Journal of Bacteriology*, 195(11), 2632-2641.
27. Hagiwara, D., Yamashino, T., & Mizuno, T. (2004). Genome-wide comparison of the His-to-Asp phosphorelay signaling components of three symbiotic genera of Rhizobia. *DNA Research*, 11(1), 57-65.

28. Patriarca, E. J., Tatè, R., & Iaccarino, M. (2002). Key role of bacterial NH4(+) metabolism in Rhizobium-plant symbiosis. *Microbiology and Molecular Biology Reviews*, 66(2), 203-222.
29. Wang, D., Xue, H., Wang, Y., Yin, R., Xie, F., & Luo, L. (2013). The Sinorhizobium meliloti ntrX gene is involved in succinoglycan production, motility, and symbiotic nodulation on alfalfa. *Applied and Environmental Microbiology*, 79(22), 7150-7159.
30. Calatrava-Morales, N., Nogales, J., Ameztoy, K., van Steenbergen, B., & Soto, M. J. (2017). The NtrY/NtrX system of Sinorhizobium meliloti GR4 regulates motility, EPS I production, and nitrogen metabolism but is dispensable for symbiotic nitrogen fixation. *Molecular Plant-Microbe Interactions*, 30(5), 566-577.
31. López, M. F., Hegel, V. A., Torres, M. J., Hidalgo-García, A., Delgado, M. J., & López-García, S. L. (2019). The Bradyrhizobium diazoefficiens two-component system NtrYX has a key role in symbiotic nitrogen fixation of soybean plants and cbb3 oxidase expression in bacteroids. *Plant and Soil*, 440(1-2), 167-183.
32. Drepper, T., Wiethaus, J., Giaourakis, D., Gross, S., Schubert, B., Vogt, M., Wiencek, Y., McEwan, A. G., & Masepohl, B. (2006). Cross-talk towards the response regulator NtrC controlling nitrogen metabolism in Rhodobacter capsulatus. *FEMS Microbiology Letters*, 258(2), 250-256.
33. Poindexter, J. S. (1964). Biological properties and classification of the Caulobacter group. *Bacteriological Reviews*, 28(3), 231-295.
34. Fröhlich, K. S., Förstner, K. U., & Gitai, Z. (2018). Post-transcriptional gene regulation by an Hfq-independent small RNA in Caulobacter crescentus. *Nucleic Acids Research*, 46(20), 10969-10982.
35. Fernández, I., Sycz, G., Goldbaum, F. A., & Carrica, M. D. C. (2018). Acidic pH triggers the phosphorylation of the response regulator NtrX in alphaproteobacteria. *PLOS ONE*, 13(4), e0194486.

36. Foulongne, V., Bourg, G., Cazevieille, C., Michaux-Charachon, S., & O'Callaghan, D. (2000). Identification of Brucella suis genes affecting intracellular survival in an in vitro human macrophage infection model by signature-tagged transposon mutagenesis. *Infection and Immunity*, 68(3), 1297-1303.
37. Atkinson, M. R., & Ninfa, A. J. (1998). Role of the GlnK signal transduction protein in the regulation of nitrogen assimilation in Escherichia coli. *Molecular Microbiology*, 29(2), 431-447.
38. Roop, R. M. 2nd, Gaines, J. M., Anderson, E. S., Caswell, C. C., & Martin, D. W. (2009). Survival of the fittest: How Brucella strains adapt to their intracellular niche in the host. *Medical Microbiology and Immunology*, 198(4), 221-238.
39. Ma, L. (2023). Application of bioinformatics in microorganisms. *Industrial Microbiology*, 53(04), 27-29.
40. Gu, N., Wang, F., Li, Y., Tang, T., Cao, C., & Shen, Y. (2022). Cell Biology and Technology. *Science in China: Chemistry*, 52(09), 1673-1684.
41. Long, S. (2024). Comprehensive analysis and targeted drug screening of AURKA oncogene in cutaneous melanoma based on bioinformatics and molecular docking technology. *Lanzhou University*.
42. Marsico, F., Burastero, O., Defelipe, L. A., et al. (2018). Multiscale approach to the activation and phosphotransfer mechanism of CpxA histidine kinase reveals a tight coupling between conformational and chemical steps. *Biochemical and Biophysical Research Communications*, 498(2), 305-312.
43. Liu, Y., Yang, X. C., Gan, J. H., et al. (2022). CB-Dock2: Improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Research*, 50(D1), 159-164.
44. Yang, X. C., Liu, Y., Gan, J. H., et al. (2022). FitDock: Protein–ligand docking by template fitting. *Briefings in Bioinformatics*, 23(3), bbac087.

45. Yang, J. Y., Roy, A., & Zhang, Y. (2013). Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, 29(20), 2588-2595.
46. Yang, J. Y., Roy, A., & Zhang, Y. (2013). BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions. *Nucleic Acids Research*, 41, D1096-D1103.
47. Koh, A., Gibbon, M. J., Kamp, M. W., et al. (2021). Conformation control of the histidine kinase BceS of Bacillus subtilis by its cognate ABC-transporter facilitates need-based activation of antibiotic resistance. *Molecular Microbiology*, 115(1), 157-174.

ACKNOWLEDGMENTS

I would like to sincerely thank Teacher Tang Ke for helping me successfully complete this paper under his careful guidance.