Insights into the substrate binding specificity of quorum-quenching acylase PvdQ

A Dissertation Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Yanyun Liu

May 2019

Insights into the substrate binding specificity of quorum-quenching acylase PvdQ

Yanyun Liu

APPROVED:

Dr. James M. Briggs, Chairman

Dr. George E. Fox

Dr. Jason A. Rosenzweig
Texas Southern University

Dr. Weihua Zhang

Dr. Dan E. Wells, Dean, College of Natural Sciences and Mathematics

Acknowledgements

I would like to express my deepest gratitude to my advisor, **Dr. James M. Briggs**, for his excellent guidance, caring and continuous support throughout my Ph.D. study and dissertation research. I still remember the first meeting with him in 2014. I was very nervous to meet him, but he was so nice and friendly and made me believe in myself. I have always been interested in computational study and really wanted to join a computational lab for my Ph.D. research, even though I didn't have a strong computational background. He wasn't concerned about my background and encouraged me to pursue my research interest, gaving me the chance to join his lab. He is very patient to guide me to learn all the knowledge needed, helps me to think critically about my ideas, and to gain deep understanding of my research. When I first came to US, my English was very bad and I can't even understand what people say and didn't know how to express myself clearly. Dr. Briggs always patiently explains the meaning of vocabularies I don't know and corrects my grammar. Moreover, during my pregnancy, Dr. Briggs helped me a lot to make my life and research easier. Without his sincere help and supportive research environment, I would never have been able to complete my research and dissertation.

I sincerely thank the rest of my committee members, **Drs. George E. Fox, Jason A. Rosenzweig, and Weihua Zhang**. The completion of this dissertation is a testament to their guidance, feedback, and valuable insights throughout the duration of my research. I am thankful to them for holding me to a high research standard and inspiring me to be a better researcher.

I would also like to thank **Dr. Mehmet Sen** who give me the chance to work with him and learn new knowledge. Collaborating with him is a valuable experience for me to expand my horizon and learn how to cooperate with other researchers to make greater success. I would like to thank him for his insightful comments and encouragement.

To my first mentor: **Dr. Xiaolian Gao**, thank you for accepting me into your lab to do experiments. Without your help, I don't know where I would be now. Your passion for science really amazes me.

I am grateful to the other members of the Dr. Briggs lab. Their generous support and care helped me stay focused on my research. Many thanks to **Dr. Iris Nira Smith**, who is the best lab mate I ever had. Thanks so much for mentoring me to do my research and help me in my personal life. You are like a big sister to me and I really appreciate your various forms of support and I am grateful to have you as my friend in my life. I would like to thank **Dr. Khushboo Singh** for her help with my research and we had a lot of fun together. For **Dr. Benjamin L. Skidmore**, there is no word can express my thanks to him. We talk, eat, and laugh together almost every weekday. With his help, my English has improved a lot and I have learned a lot of American culture from him. He is like my big brother to help me go through my difficulties even though he is younger than me :). For **Dr. John W. Craft**, I would like to thank him for his help when I first came here and I enjoyed talking to him. For **Sladja Maric**, I would like to thank her for the suggestions during our lab meetings and it is so great to have her as part of our lab family. For Zahra Mazhar and Jasper Nyandoto, thanks for helping to prepare my defense presentation. I would like to thank **Dr. Guedmiller Souza de Oliveira** for solving many problems for my project and I am so glad we cooperated on one paper together. Thank you all.

Many thanks to my colleagues, **Fei Yuan, Haopeng Yang, Celise Robertson, Priyanka Srivastava, Pragya Manandhar, Shivangi Srivastava, Nada Béjar, Savini Wellappuliarachchi, …** for their generous help and support during these years. They are great friends.

To my family: **Mom** and **Dad**, thanks so much for bringing me into this world and always support me in my study. I know deep in heart that you don't want your little girl be away from you. I would like to thank you for always understanding and encouraging me. Thanks so much for taking good care of my baby and me. To my **sister** and **brother**, thank you for always taking care of me and our parents when I am not around. I would like to thank my **mother-in-law** for helping me take good care of my baby. To my husband, **Dr. Ning Cao**, there are no words strong enough to express my deepest love and gratitude to you. Without your support, I would never be able to reach at this point in my life. I love you and looking forward to this exciting and new chapter of our life. Finally, to our dearest daughter **Sophia**, thanks so much for choosing us as your parents. You are the best thing ever happen to us. You bring so much laugh and joy to our life. Mom and Dad love you forever.

Insights into the substrate binding specificity of quorum-quenching acylase PvdQ

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Biology and Biochemistry

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Yanyun Liu

May 2019

Abstract

The *N*-acyl homoserine lactone acylase, PvdQ, from human opportunistic pathogen *Pseudomonas aeruginosa* is a quorum-quenching enzyme that can hydrolyze the amide bond of the quorum sensing signaling *N*-acyl homoserine lactones (AHLs) thereby degrading the signaling molecules, inhibiting the biofilm formation and reducing virulence gene expression. Previous studies demonstrated that PvdQ has different preferences for AHLs with different acyl chain lengths and substituents. However, the substrate binding specificity determinants of PvdQ with different bacterial ligands remain unknown and unintuitive. Elucidation of these determinants can lead to mutants with efficiency and broader substrate promiscuity.

To investigate this question, a computational study was carried out combining multiple molecular docking methods, molecular dynamics (MD) simulations, residue interaction network analysis, and binding free energy calculations. The main findings are: firstly, results from pKa predictions support the observation that the pKa of the N-terminus of Serβ1 was depressed due to the surrounding residues. Multiple molecular docking studies provided information about PvdQ binding modes and binding affinities. Secondly, analysis of the protein dynamic fingerprint of each complex from MD simulations demonstrated that binding of C12-homoserine lactone (C12-HSL) ligand reduced the global motion of the complex and maintained the correct arrangement of the catalytic site. Further, the residue interaction network analysis of each system illustrated that there are more communication contacts and pathways between the residues in the C12-HSL complex as compared to other complexes. The binding of the C12-HSL ligand facilitates

structural communication between the two knobs and the active site. The binding of other ligands tends to impair these specific communication pathways, leading to a catalytically inefficient state. Finally, simulation results from free energy landscape and binding free energy analysis revealed that the C12-HSL ligand has the most favorable binding free energy and greater stability than the less favored ligands. Each of the following residues: Serβ1, Hisβ23, Pheβ24, Metβ30, Pheβ32, Leuβ50, Asnβ57, Thrβ69, Valβ70, Trpβ162, Trpβ186, Asnβ269, Argβ297 and Leuα146, play different roles in substrate binding specificity. This is the first computational study that provides molecular information for structure-dynamic-function relationships of PvdQ with different bacterial ligands and demonstrates determinants of substrate binding specificity.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AHL	Acyl Homoserine Lactone
HSL	Homoserine Lactone
MD	Molecular Dynamics
P. aeruginosa	*Pseudomonas aeruginosa*
B. cenocepacia	*Burkholderia cenocepacia*
CF	Cystic Fibrosis
QS	Quorum Sensing
LPS	Lipopolysaccharide
PCR	Polymerase Chain Reaction
Bcc	*Burkholderia cepacia* complex
AI	Autoinducer
QQ	Quorum Quenching
SAM	S-adenosyl-methionine
acyl-ACP	acyl-acyl-carrier-protein
MTAN	5′-methylthioadenosine/S-adenosylhomocysteine nucleosidase
UHBD	University of Houston Brownian Dynamics
SP	Standard-Precision
XP	Extra Precision
vdW	van de Waals
GROMACS	GROningen MAchine for Chemical Simulations
QM	Quantum Mechanics

CM	Classical Mechanics
IFD	Induced Fit Docking
RMSD	Root Mean Square Deviation
RMSF	Root Mean Square Fluctuation
Rg	Radius of gyration
CHARMM	Chemistry at HARvard Macromolecular Mechanics
PCA	Principal Component Analysis
FEL	Free Energy Landscape
RIN	Residue Interaction Network
PSN	Protein structure network
ENM-NMA	Elastic Network Model-Normal Model Analysis
MM-PBSA	Molecular Mechanics-Poisson-Boltzmann Surface Area
MM-GBSA	molecular Mechanics-Generalized Born Surface Area
SASA	Solvent Accessible Surface Area

Chapter 1 Introduction

Pseudomonas aeruginosa (*P. aeruginosa*) and *Burkholderia cenocepacia* (*B. cenocepacia*) are prevalent opportunistic pathogens that are the primary pathogens of concern for cystic fibrosis (CF) patients. Both, individually or in co-infection, can cause severe chronic pulmonary infections in immunocompromised and CF patients, with high infection and fatality rates. When co-inhabiting the same environment, both use quorum sensing (QS) systems and can form mixed biofilms in the lungs of CF patients. The interspecies cooperation of *P. aeruginosa* and *B. cenocepacia* during an infection has emerged as a major new challenge due to high fatality and resistance to most common antibiotics in patients. Therefore, it is important to understand the basic knowledge about these bacteria and develop new strategies to target these two bacteria.

In this chapter, the background information of these two bacteria, quorum sensing, and quorum quenching enzymes are introduced along with the hypothesis and scope for this study.

1.1 Pseudomonas aeruginosa

1.1.1 Overview

P. aeruginosa is a ubiquitous, rod-shaped, Gram-negative, aerobic, motile bacterium. *P. aeruginosa* is a multidrug resistant opportunistic pathogen capable of causing serious infection.[1-3] It is one of the most common causes of infection in humans and the infection is riskier for patients who are immunocompromised, particularly transplant recipients, neutropenic patients, and patients with cancer, severe burn woods, and HIV. *P.*

aeruginosa is responsible for acute respiratory infections in immunocompromised patients, chronic respiratory infection in CF patients, and excessive mortality in ventilator-associated pneumonia patients.[1,4-8] More specifically *P. aeruginosa* is responsible for 26% of ventilator-associated pneumonia,[9] 12% of catheter-associated urinary tract infections, and 15.4% of surgical site infections.[10,11] Finally, it is also a major cause of morbidity and mortality in CF patients.[12,13]

Even though various strategies and therapeutic solutions have been proposed to deal with *P. aeruginosa*,[2,14-20] the problems of seriousness of infection, high fatality rates, and increasing resistance to common antibiotics still persist, and the need for novel therapeutic options remains.

1.1.2 Pathogenesis

The pathogenicity of *P. aeruginosa* is mainly caused by many different bacterial virulence factors (**Figure 1.1**) and its genetic flexibility that allows it to adapt to hostile environments. The precondition of *P. aeruginosa* infection typically requires a loss of first-line defense by the host, such as breakdown of the skin or mucosal barriers or an impaired immune system. In general, *P. aeruginosa* infection consist of three stages: (i). attachment and colonization, (ii). local invasion, and (iii). systemic disease. For each of these stages, a progression of virulence factors is involved.

Virulence factors used in attachment and colonization include: flagella, pili, and lipopolysaccharide (LPS). Flagella are involved in motility, adherence, and invasiveness.[21] Type IV pili mediate the adherence to the epithelium. In addition, LPS also contributes to the bacterial adhesion by binding to host asialo-GM1. These

appendages are mainly responsible for the adhesion to epithelial cells and enable colonization of the respiratory epithelium.[1]

After colonization, several extracellular products of *P. aeruginosa* can cause severe tissue damage, bloodstream invasion, and dissemination.[14] Exotoxin A plays a major role in local tissue damage, bacterial invasion, and possibly immunosuppression.[22] Phospholipase C and rhamnolipid may break down lipids and lecithin, and disrupts lung surfactants. Alkaline protease degrades immune system components, such as complement factors.[23,24] Two elastases, LasB and LasA, play important roles in degrading elastin, disrupting membranes, impairing monocyte chemotaxis, and degrading complement proteins.[25]

Upon attachment, a powerful tool called the type III secretion system is activated to inject effector proteins, such as ExoS, ExoT, ExoU and ExoY, through a syringe-like apparatus into the eukaryotic cytoplasm.[1] All four effector proteins play important roles in the cytotoxicity and cause invasion and promote bacterial dissemination. [23]

The mechanisms of *P. aeruginosa* antibiotic resistance include: low outer membrane permeability,[14,26] multidrug efflux pumps, and the production of antibiotic modifying enzymes.[14,27,28] *P. aeruginosa* can also evade clearance by losing immunogenic features such as pili and flagella.[1] Additionally, *P. aeruginosa* can form biofilms to protect itself from host immune cell clearance. Biofilms promote persistent chronic infections. Moreover, *P. aeruginosa* also has regulating systems called QS to coordinate the bacterial population in adapting to the changing environment.

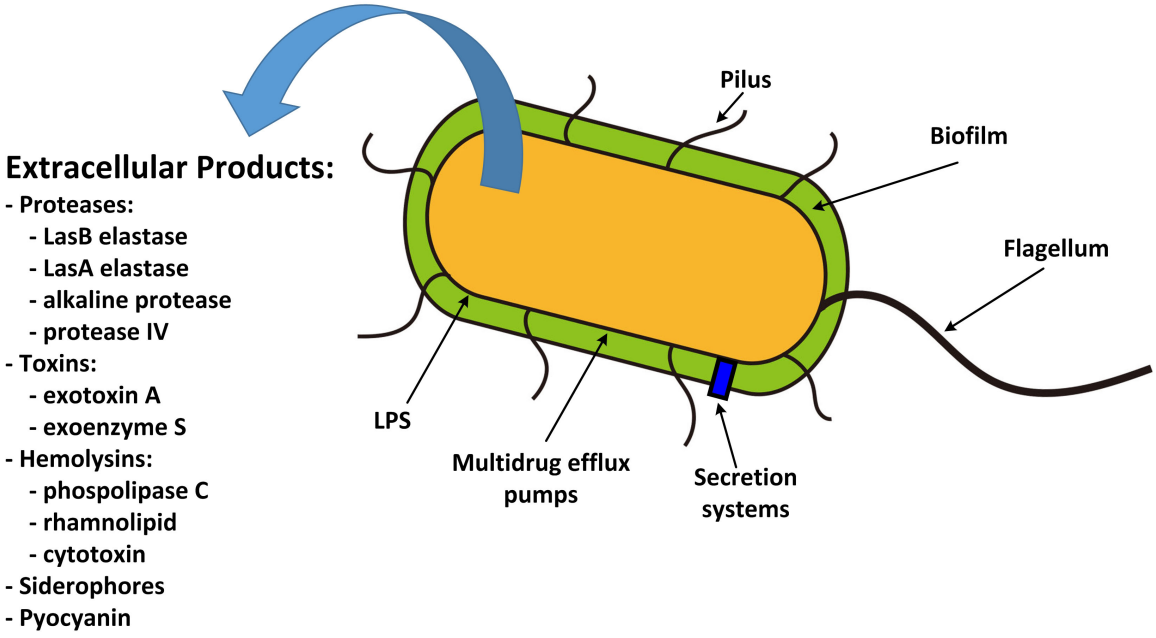

Figure 1.1. Virulence factors of *P. aeruginosa*.

(Adatped from Sorger-Domenigg T.[14])

1.1.3 Diagnosis

In the healthcare setting, *P. aeruginosa* can be spread through improper hygiene or via contaminated medical equipment. *P. aeruginosa* can infect the blood, lungs, urinary tract, wounds, ear, eyes, and skin. Bloodstream infections present with symptoms such as: fever and chills, muscle and joint pain and stiffness, body aches, rapid pulse and breathing, fatigue, nausea and vomiting, diarrhea, and decreased urination. Respiratory tract infections present with symptoms of coughing, congestion, fever and chills. Urinary tract infections present with painful urination, odor in urine, cloudy or bloody urine. Wound infections present wtih inflammation, green pus or discharge leakage from wound while ear infections causing hearing loss, ear pain, ear discharge, itching inside the ear, dizziness, and disorientation. Eye infections can cause eye pain, redness, swelling, and

impaired vision while skin infections typically present with rash, which can include pimples filled with pus.

Early diagnosis is critical for treating *P. aeruginosa* infections. The basic identification of *P. aeruginosa* is to use conventional methods, indirect molecular strategies and direct molecular strategies including polymerase chain reaction (PCR) and sequence analysis. Conventional culture methods are still the most commonly used methods in clinical practice. *P. aeruginosa* is identified based on the biological characteristics of the bacterium under certain culture conditions, such as Gram-negative/Gram-positive status, or the activities of bacterial molecules such as oxidase, acetamidase, arginine dihydrolase, glutamate utilization, proteolytic activity, nitrate utilization, lipolytic activity, and pyocyanin pigment formation.[29,30] Several automated identification systems are available for *P. aeruginosa* identification, such as the Vitek2[31] and Phoenix 100 systems.[32] In recent years, numerous modern detection approaches, such as flow cytometry,[33] immunological detection,[34] and molecular biology-based detections,[35] have been developed to establish a rapid and sensitive detection of *P. aeruginosa* infections, especially for early detections.[29] The PCR-based identification methods have also becoming popular for rapidly detecting *P. aeruginosa*.[36-41] Conventional PCR detection targeting the 16S rRNA gene has been established to identify *P. aeruginosa*.[36] The loop-mediated isothermal amplification based PCR method has also been developed to avoid the thermocycling for amplification.[42] Since early diagnosis of *P. aeruginosa* is critical to effectively treat the infections, developing affordable, rapid, and accurate identification methods is of significant importance.

1.1.4 Treatment

Different strains are resistant to different antibiotics. Due to this wide range of resistance, finding the proper therapies to treat *P. aeruginosa* is challenging. *P. aeruginosa* is naturally resistant to many antibiotics, such as tetracyclines and benzylpenicillin due to low outer membrane permeability and has increasing resistance to new antibiotics and drugs.[43] Moreover, selection can occur during treatment and trigger the development of bacterial resistance.[44,45]

In general, there are several groups of antimicrobials that can be used to treat *P. aeruginosa*.[46] (i). Penicillins (such as β-lactams, ticarcillin, tazobactam, and piperacillin) and cephalosporins (such as cefepime and ceftazidime) have been used to treat *P. aeruginosa*.[47-49] (ii). Carbapenems (such as meropenem, doripenem, ertapenem, and imipenem) are generally the main drugs used in the treatment of *P. aeruginosa*.[50] (iii). Aminoglycosides (such as amikacin, netilmicin, tobramycin, and gentamicin) can also be used in combination treatment of *P. aeruginosa*.[51] (iv). Monobactam (such as aztreonam) has been used to treat *P. aeruginosa*, especially for urinary tract infections, intra-abdominal, soft tissue infections, and pelvic infections and pneumonia.[46,52,53] (v). Colistin has also shown efficacy in treating *P. aeruginosa*.[54-57] In general, a drug cocktail of several antibiotics that are effective against most strains and have different modes of actions should be used.[58] Even though tobramycin can effectively eradicate early *P. aeruginosa*, it has been found that *P. aeruginosa* has become increasingly resistant to tobramycin.[2,59-62] *P. aeruginosa* shows limit resistance to colistin in cystic fibrosis. However, colistin has potential concerns for neurotoxicity and nephrotoxicity that need to be monitored.[63]

Even though some of the above antibiotics are effective against *P. aeruginosa*, these antibiotics are not effective against all strains. Selecting the right antibiotic requires a clinical isolate from a patient which can be typed to determine the most effective treatment. Early treatment with suitable drugs is significant to optimize the clinical outcome and numerous factors should be considered.[64] However, it is difficult to select the best antipseudomonal due to the increasing resistance of *P. aeruginosa*, which has become a serious problem in recent years.

1.2 Burkholderia cenocepacia

1.2.1 Overview

B. cecocepacia belong to the Burkholderia cepacian complex (*Bcc*), which are Gram-negative, aerobic, motile, rod-shaped bacteria existing in diverse environments such as soil, water, plants, animals, and human.[65] *Bcc* are a group of closely related opportunistic organisms that are more infective to patients with chronic granulomatous disease, who are immunocompromised, or have pre-existing lung conditions, such as CF. The *Bcc* consists of at least 17 species, including *B. cepacia*, *B. multivorans*, and *B. cenocepacia* etc.[66-69] *Bcc* is the second leading cause of chronic lung infection behind *P. aeruginosa*. As an important pathogen in CF pulmonary infections, *Bcc* often infect CF patients in the late process of the disease when they are already chronically infected with *P. aeruginosa*. Such co-infection causes variable and unpredictable consequences in clinic, and can range from an asymptomatic carriage to the fatal pneumonia. The so-called “cepacia syndrome” is a condition manifesting in high fever, leukocytosis and progressive respiratory failure.[70] Among all the *Bcc*, *B. cenocepacia* has been one of the most

prevalent in CF. The main difficulty in dealing with *B. cenocepacia* is their extraordinary resistance to antimicrobial agents. One important factor of the infection process is its biofilm formation that can withstand host immune responses and provides increased resistance to antibiotics. Moreover, along with *P. aeruginosa*, mixed biofilms can be formed, which is more problematic in clinical settings.

1.2.2 Pathogenesis

B. cenocepacia possess numerous pathogenic mechanisms that enable them to invade, hide, and survive in their natural hosts[71-77] including flagellum,[71] extracellular proteases,[76] siderophore production,[72] a type III secretion,[75] biofilm formation,[73] and a QS mechanism (**Figure 1.2**).[74]

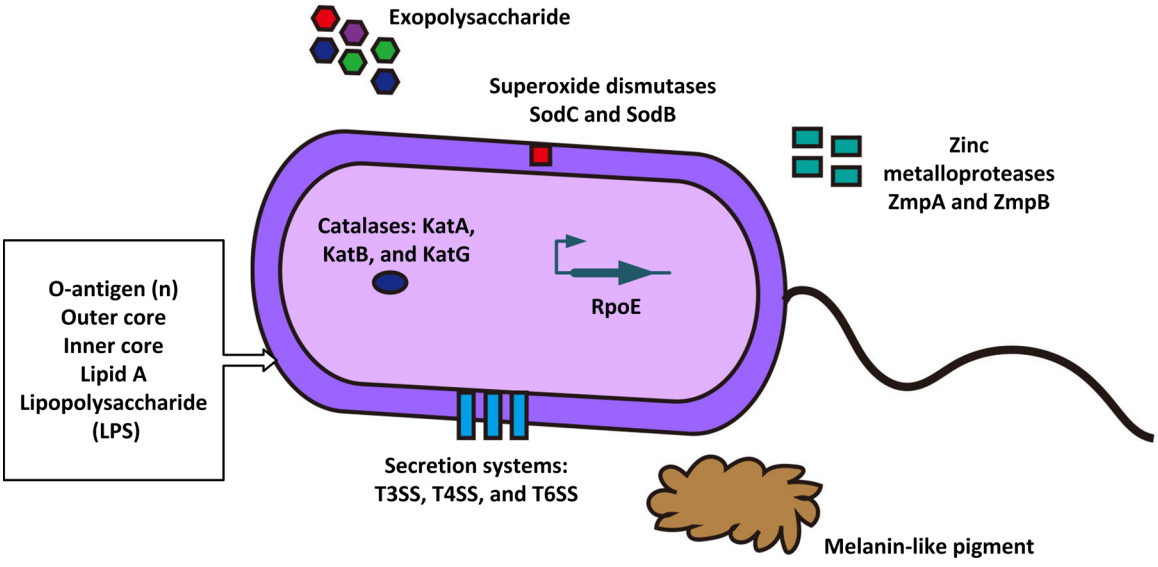

Figure 1.2 Important virulence factors for *Bcc* survival in interactions with host phagocytes. (Adapted from Porter L. A.[77])

The first step of the infection is the adhesion of the *B. cenocepacia* to the epithelial surface of the host cell via protein and glycolipid receptors as well as secretory mucins.

B. cenocepacia possess flagellum that contribute to bacterial motility and adhesion and enable the organism to invade host cells.[71,78] In addition to the above-mentioned adhesion and invasion processes, *Bcc* have many virulence factors, such as proteases, lipases, and LPS, to enhance the pathogenicity in epithelial cells. *B. cenocepacia* can produce extracellular proteases, such as zinc metalloproteases (i.e., ZmpA and ZmpB).[76-79] ZmpB has been found to be able to cleave immunoglobulins, transferrin, and lactoferrins. The lipase produced by *B. cenocepacia* plays an important role in invasion of lung epithelial cells. The LPS located in the outer membrane of *B. cenocepacia* plays important roles in contribution to antimicrobial peptide resistance and promotion of a potent pro-inflammatory response.[80] The O-antigen portion within the LPS molecular structure plays an important role for resistance to serum-mediated killing[81] and prevents bacterial binding to epithelial cells and phagocytosis by macrophages.[82] *B. cenocepacia* can produce different siderophores to acquire iron under conditions of iron depletion.[83] Similar with *P. aeruginosa*, *B. cenocepacia* also possess the Type III secretion system to directly inject proteins from the bacterial cytoplasm into the host cell cytoplasm.

B. cenocepacia possess various mechanisms to resist oxidative stress with catalase, peroxidase, and superoxide dismutase activities.[84] *B. cenocepacia* can produce biofilms, a complex, multicellular bacterial communities to protect bacterial from antibiotics and the host immune system.[85] In CF lungs, *B. cenocepacia* can also produce mixed biofilms with *P. aeruginosa* and communicate with *P. aeruginosa* via QS systems.[86,87] The QS mechanism is for cell-to-cell communication to regulate protease production, siderophore synthesis, the type III secretion system, motility and biofilm formation.[88] QS enables bacteria to coordinate the behavior of their community in a cell-density-dependent

manner by controlling the expression of the virulence-related genes.

1.2.3 Diagnosis

Due to the seriousness of the *B. cenocepacia* infection, laboratory diagnosis should be carried out as accurately and early as possible. However, in practice, isolation and diagnosis of *B. cenocepacia* is complicated and difficult, especially for routine microbiological laboratories. In general, the *B. cenocepacia* can be identified by a combination of selective media and biochemical analysis in the clinical laboratory. These bacteria can be difficult to isolate since other bacteria from the patient sample may overgrown them. The selective media is based on the *B. cenocepacia*'s high intrinsic resistance while suppresses the growth of non-*Bcc* organisms. *B. cenocepacia* can be identified phenotypically with conventional culture with selective media using samples from blood, sputum, cerebrospinal fluid or other clinical specimens. A definitive diagnosis includes features like oxidase-positive, polymyxin-resistant, and gentamicin-resistant. PCR-based assays have also been developed to amplify the 16S rRNA gene.[89-91] After 16S rRNA amplification, restriction enzyme mediated fragmentation can generate a specific restriction fragment length polymorphism (RFLP) banding pattern to identify the *B. cenocepacia*.[92] RFLP method based on amplifying the recA gene has also been used to identify the *B. cenocepacia*.[93] Other identification methods include amplified fragment length polymorphism (AFLP) fingerprint typing,[94] sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) of whole cell proteins, and whole cell fatty acid analysis.[67]

1.2.4 Treatment

The antimicrobial resistance of *B. cenocepacia* is based on mechanisms of cell wall impermeability, enzymatic inactivation, alteration of drug targets, and active efflux pumps.[95] Antibiotics, such as piperacllin, azlocillin, cephalosporins, meropenem, ciprofloxacin, minocycline, chloramphenicol, carbapenems, fluoroquinolones, ceftazidime, semisynthetic penicillins, and trimethoprim-sulfamethoxazole have been used to treat *Bcc*.[96-99] In clinical practice, combinations of different antibiotics are generally more effective, such as combinations including meropenem, co-trimoxazole, chloramphenicol, and tetracyclines[97] or a four-drug combination of chloramphenicol, doxycycline, and trimethoprim-sulfamethoxazole.[100] Inhalational immunotherapy, which has emerged as important to manage chronic lung infections, can also be used to acute *B. cenocepacia* infection by combining with other antibiotics.[101] *B. cenocepacia* as well as the other strains in *Bcc* are difficult to eradicate and the treatment should be made on a case-by-case basis.

1.3 Quorum Sensing

1.3.1 Quorum sensing mechanism

As mentioned above, *P. aeruginosa* and *B. cenocepacia* are both highly communicative organisms with the purpose of coordinating their behavior, which is referred to as QS. QS is a cell to cell signaling mechanism that enables the bacterium to regulate cell density and coordinate virulence factors through small signaling molecules, controlling antibiotics production, DNA exchange, motility, and biofilm production. Those signaling

molecules released by bacteria in the extracellular medium are called autoinducers and work as mediators of QS. As shown in **Figure 1.3**, typically the signaling molecule is N-acyl homoserine lactone, which is synthesized by an autoinducer synthase, and recognized by an autoinducer receptor. When the bacterial population increases sufficiently, the autoinducer concentrations will increase and exceed a particular threshold level. These signaling molecules are internalized into the cell and bind the transcription activators, which activate expression of particular genes. This includes genes responsible for biofilm formation, various virulence factors, and other group behaviors.

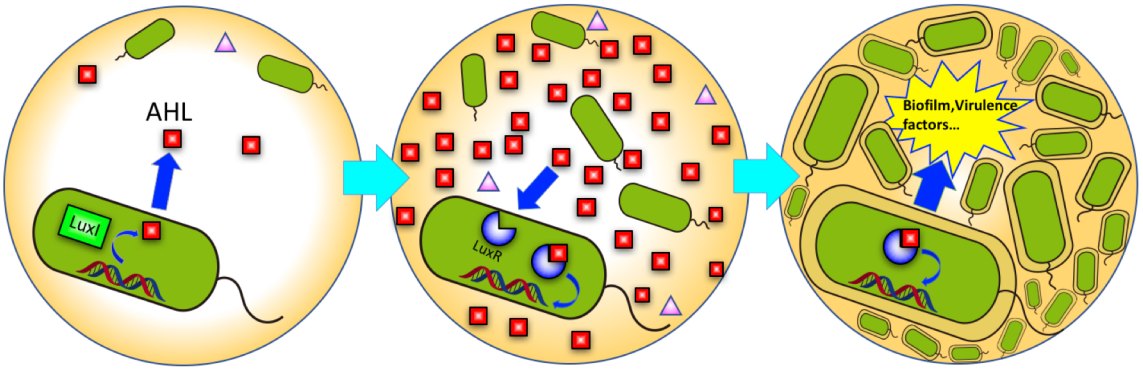

Figure 1.3 Quorum sensing mechanism. It includes three steps: synthesis, recognition, and response. Bacteria synthesize the quorum sensing molecules, which diffuse into the surrounding environment at low cell densities. However, when the local population increases in density and the concentration of the signaling molecules exceeds a particular threshold level, these molecules are internalized into the cell and bind to the cognate receptor; then activate a particular set of genes.

1.3.2 Quorum sensing signal molecules

Different bacteria produce different signal molecules. As indicated in **Figure 1.4**, there are generally three categories of signal molecules: Gram-negative bacteria primarily use AHLs, known as Autoinducer-1 (AI-1); Gram-positive bacteria that mainly use

autoinducing peptides (AIPs); A furanosyl borate diester, also known as Autoinducer-2 (AI-2), is the only species-nonspecific autoinducer and can be seen as a universal QS signal.[102] AI-2 is commonly available in both Gram-negative and Gram-positive bacteria for intraspecies and interspecies communication.[103]

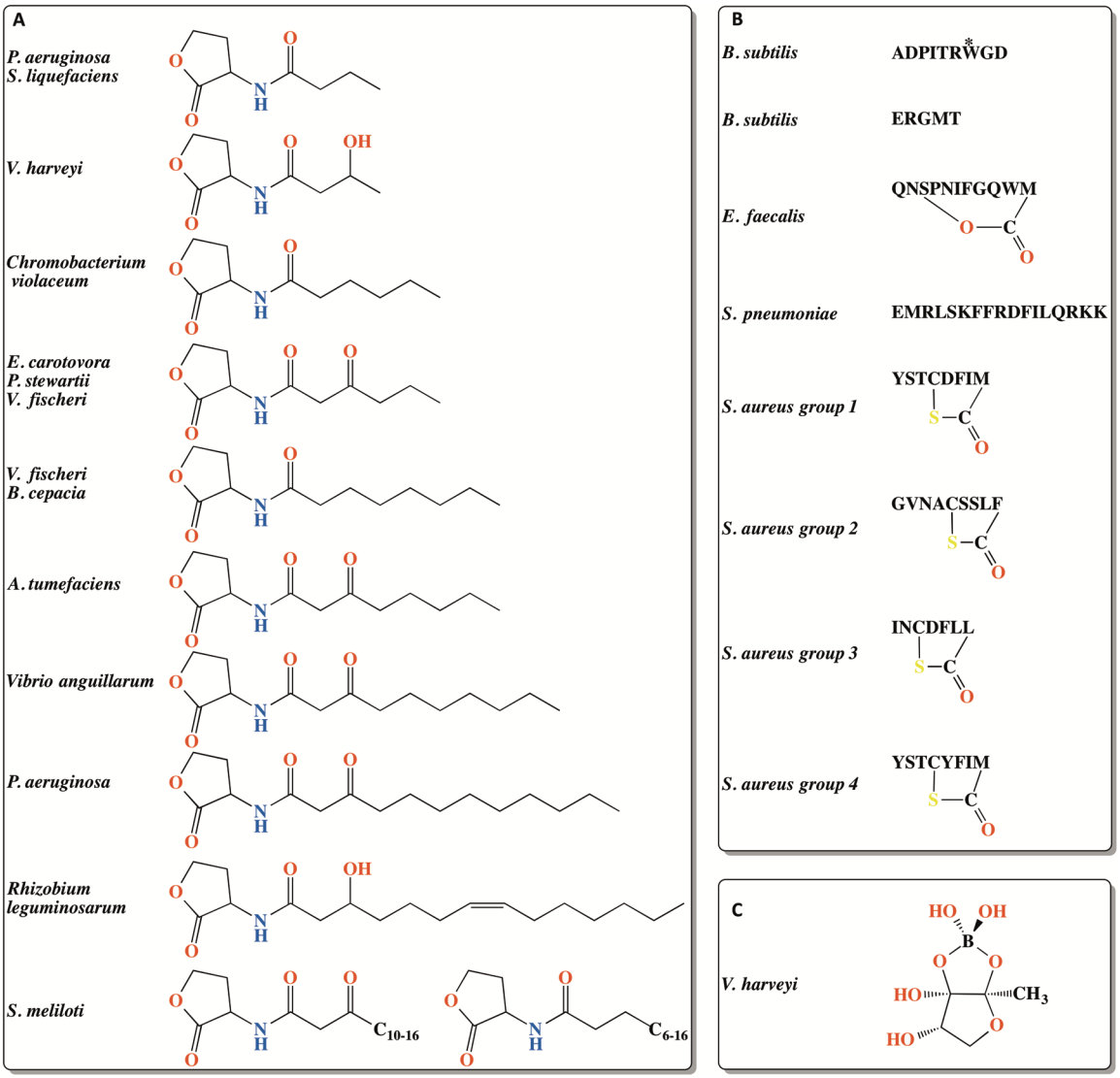

Figure 1.4 Molecular structures of different autoinducers. (A) Different AHLs with the corresponding bacteria. (B) AIPs with corresponding bacteria. The asterisk represents an isoprenyl group. (C) AI-2 with corresponding bacteria.

[Adapted from Taga M. E.[102]]

AHLs are diffusible, low molecular weight molecules and all of which share a common homoserine lactone head group with different acyl side chains. The differences in side chain length, saturation and oxidation at the C_3 position results in different AHLs. Because of amphipathic nature, AHLs can diffuse freely in and out of the cell in most conditions, which means AHL concentration in the extracellular environment increases when the cell density increases. However, active transport of AHLs also happens.[104] Short chain AHLs can diffuse freely and efficiently across the cell envelope while long chain AHLs diffuse slowly due to the interactions with the lipophilic cytoplasmic membrane.[105] In addition, it has been found that long chain AHLs can be exported by efflux pumps in certain bacterial strains.[104,106,107]

Unlike AHLs, AIPs cannot permeate the cell membrane and require specialized transporters for active transport. Detection of the AIPs can occur either at the surface of the membrane or intracellularly. For some cases, AIPs cannot directly bind to a transcriptional activator. They need to bind to a receptor on the surface of the membrane, which uses a phosphorylation based relay system to eventually activate or repress the QS target genes.

In general, AI-2 can freely diffuse out of both Gram-negative and Gram-positive bacteria. Detection of AI-2 can occur either extracellularly or intracellularly depending on the bacterium. The synthesis of AI-2 is executed by LuxS, which is an enzyme that cleaves S-ribosyl-L-homocysteine to produce L-homocysteine and 4,5-dihydroxy-2,3-pentanedione, upon which the latter undergoes a spontaneous intramolecular cyclization and hydration to generate AI-2.[108]

QS provides stimulus for unique and various cellular responses along with detection

and/or protection from competing microbial communities. Examples of species-specific and interspecies QS exist. In Gram-negative bacteria, species-specific QS is mediated by AHLs with different acyl chain length and substitution. In Gram-positive bacteria, species-specific QS is mainly mediated by small autoinducer peptides.[109] Even though the autoinducers are usually specific to its cognate receptor, nonspecific signaling or cross-talk between different QS systems can also occur. For example, interaction of *P. aeruginosa* and *B. cenocepacia* in CF patients takes place: *B. cenocepacia* can detect the AHL signals produced by *P. aeruginosa* to activate its QS system even though *P. aeruginosa* cannot directly detect *B. cenocepacia*'s signals.[110-111] For interspecies communication, autoinducer-2, a furanosyl borate diester, has been identified. QS not only regulates intraspecies survival and differentiation in bacterial communities, but is also responsible for interspecies communication between symbionts and competitors. Interactions between different Gram-negative and Gram-positive bacteria affect bacterial persistence and clinical outcomes. For example, *S. aureus* is negatively associated with subsequent *P. aeruginosa* in CF patients. *P. aeruginosa* is negatively associated with subsequent *B. cepacia* complex. *P. aeruginosa* and *B. cepacia* complex were reciprocally and positively associated with infection by *Aspergillus* species. *P. aeruginosa* and *B. cepacia* complex may inhibit other bacteria through decreasing airway biodiversity. [112] Moreover, the AiiA enzyme is an AHL lactonase from Gram-positive bacteria *Bacillus thuringiensis*, which can hydrolyze the ester bond of the homoserine lactone ring group. This enzyme exhibits different enzymatic activity toward different AHLs from Gram-negative bacteria to inhibit their "group behavior".[113] Different bacterial interaction and competition may alter coevolution and the outcomes of treatment plans.

1.3.3 Quorum sensing role in virulence and biofilm

Virulence factors are expressed by bacteria during pathogenesis and increase survival rate. QS plays an important role in regulating processes like bioluminescence, competence, biofilm formation, sporulation, antibiotic production, and virulence factor secretion.[114] Expression of virulence genes is regulated by QS in most cases and different QS systems affect virulence in different ways. In general, QS systems follow three basic steps: 1). Bacterial community produces autoinducers and the concentration of autoinducers reach a particular threshold when cell density is high, enabling autoinducer detection and response. 2). Autoinducers are recognized by cognate receptors that are located in the membrane or cytoplasm. 3). The response is to activate specific virulence gene expression. For example, in *P. aeruginosa*, QS regulates the expression of many virulence genes, including phenazine production, elastase, exotoxin A, alkaline protease, rhamnolipids, superoxide dismutases, lectins, and biofilm formation.[115] The roles of different virulence have been discussed in **Section 1.1.2**.

As mentioned earlier, bacteria can enclose themselves in a self-produced polymeric matrix and adherent to an inert or living surface, which is called biofilm. Biofilm acts as a protective coating to make the bacteria difficult to eliminate and helps them to survive in harsh environments. The mechanism of biofilm formation has five stages as shown in **Figure 1.5**. The first stage is initial attachment, where the bacteria attach onto a surface and begin to proliferate in a favorable environment. When a specific concentration of bacteria is obtained, the QS mechanism is triggered to regulate the production of extracellular polymeric substances. They form a sticky matrix causing irreversible attachment. As the colony continues to grow, the early structure of the biofilm is formed,

which contains channels in order to allow the exchange of nutriments and waste products. Then the architecture of the biofilm matures. Finally, dispersion of single cells from the biofilm occurs and planktonic bacteria leave the colony and try to find other places for new colonies.[116] QS plays an important role in the various stages of biofilm development, including the control of bacterial motility, regulating genes responsible for surface attachment and the synthesis of matrix components. It has been found that the biofilm maturation process is facilitated by controlling the swarming motility of the cells by QS.[117] Moreover, formation of biofilms needs extracellular DNA, which depends on the QS system.[118] Biofilm dispersion is also found to be regulated by the QS mechanism. Therefore, the QS system is essential for cellular aggregation, adhesion, biofilm formation, and biofilm dispersion.[119]

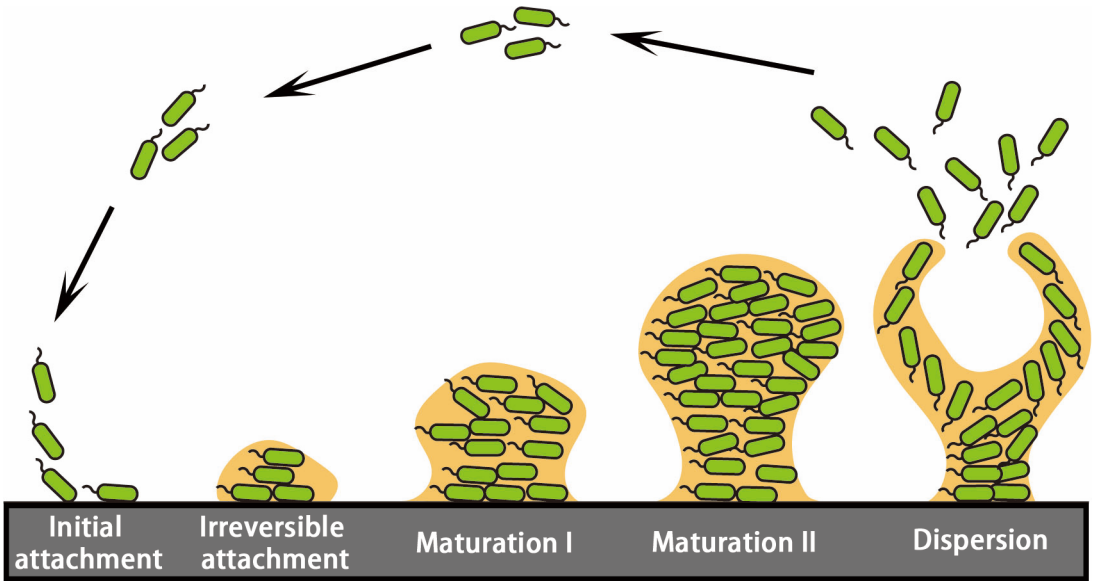

Figure 1.5 Five stages of biofilm development. First, bacteria attach onto a surface and begin to proliferate; then form a sticky matrix causing irreversible attachment. The early structure of biofilm is formed due to the colony continuing to grow; and then the architecture of the biofilm matures. Finally, dispersion of single cells occurs and planktonic bacteria leave the colony and find other new places.

[Adapted from Monroe D.[120]]

1.3.4 Quorum sensing system in *P. aeruginosa* and *B. cenocepacia*

P. aeruginosa uses three QS signal systems, i.e., Las, Rh1 and *P. aeruginosa* quinolone signal (PQS) systems (**Figure 1.6)**.[121,122] The Las and Rh1 systems are hierarchically ordered and both are AHL-dependent QS systems. Each of them is comprised of a synthase and transcriptional activator.[123] Two acyl homoserine lactone signals and one quinolone based signal are used by *P. aeruginosa*. The signal produced by the AHL systems increases as the bacterial population increases. When the signal reaches an intracellular threshold concentration, it will bind to the transcriptional regulator and regulate gene expression accordingly.[123] In the Las system, signal synthase LasI produces the signal N-3-oxo-dodecanoyl homoserine lactone (3-oxo-C12-HSL), which can bind to the transcriptional activator LasR.[124,125] Similar, in the Rhl system, signal synthase RhlI produces the signal N-butyryl homoserine lactone (C4-HSL), which can bind to the transcriptional activator RhlR.[126-128] The three systems are arranged in a hierarchy: the Las system positively regulates the Rhl system, the Las and Rh1 systems can also control the PQS synthesis, whereas PQS can stimulate the expression of the Rh1 system. Together, the three systems regulate gene expression and modulate a range of virulence factors.

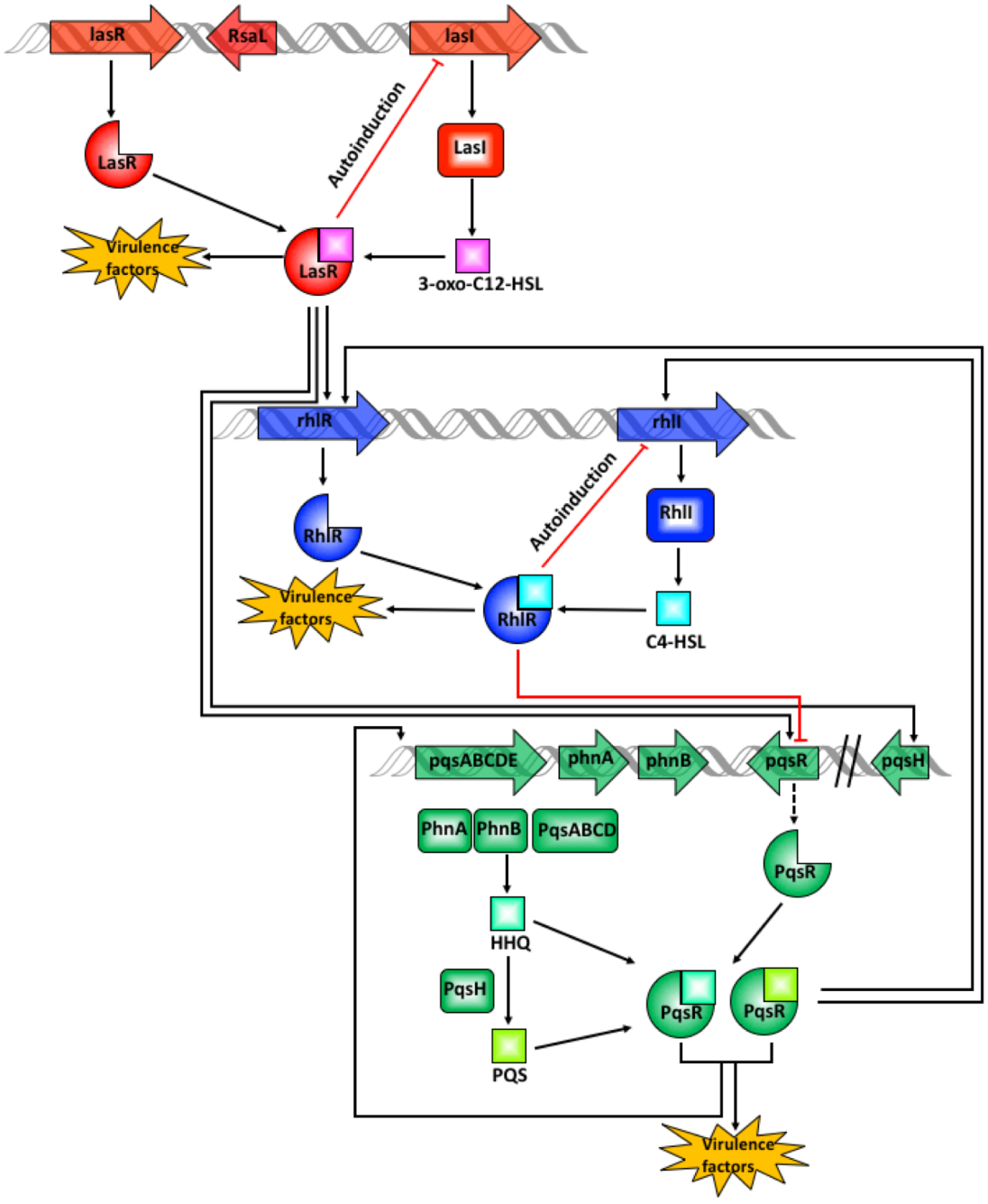

Figure 1.6 Quorum sensing in *P. aeruginosa*. The hierarchical organization of the three QS systems (Las, Rhl, and PQS systems) in *P. aeruginosa*. The Las system regulates the Rh1 system, the Las and Rh1 systems also control the PQS synthesis, while PQS can stimulate the Rh1 system.

(Adapted from Papaioannou E.[122])

The QS system in *B. cenocepacia* consists of AHL synthase CepI and AHL-dependent

transcriptional regulatory protein CepR (**Figure 1.7**).[129] CepI synthesize N-octanoyl-homoserine lactone (C8-HSL) and minor amounts of N-hexanoyl-homoserine lactone (C6-HSL) as a by-product.[130,131] When the signaling molecule reaches a particular threshold concentration, the C8-HSL binds to the cognate CepR receptor. This CepR/C8-HSL binding can activate target genes with various functions, such as biofilm formation, swarming motility, siderophores, extracellular proteases, other virulence factors, etc. CepR can control the CepI expression, likely by binding to a *lux* box-like region, which overlaps part of putative CepI promoter. Auto-regulated CepR can also negatively control itself.

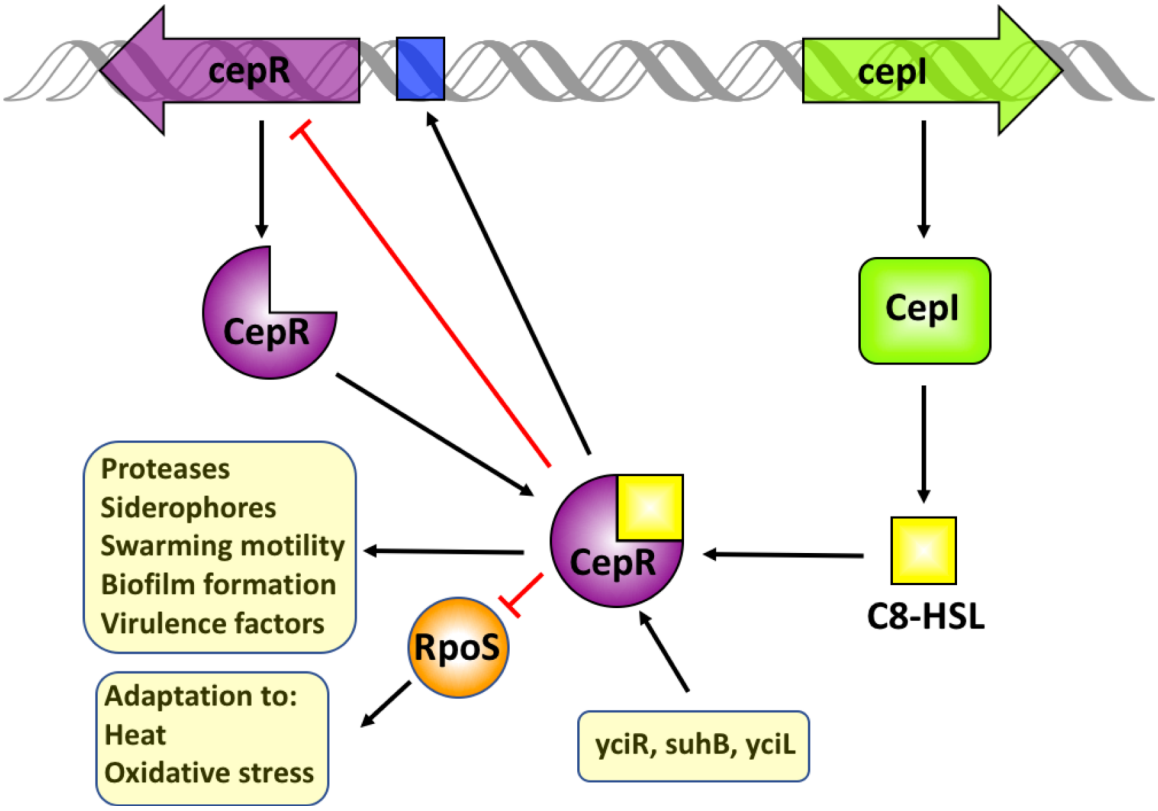

Figure 1.7 The CepI/CepR quorum sensing system of the *Bcc*. CepI synthesizes C8-HSL, which is recognized by the CepR receptor to activate target gene expression. (Adapted from Eberl L.[88])

B. cenocepacia and *P. aeruginosa* can communicate with each other and form mixed biofilms in lung infected patients. *B. cenocepacia* is able to synthesize the putative intergenus signal 2-heptyl-4(1H)-quinolone(HHQ), which can be processed into a QS signal by *P. aeruginosa*.[110] *B. cenocepacia* is able to perceive the AHL signals from *P. aeruginosa*. Since both of the two strains use AHL to mediate quorum sensing to control biofilm formation and virulence genes expression, the target of the AHL signal molecules opens up a new way to inhibit bacterial infections.

1.4 Quorum quenching enzymes

1.4.1 Quorum quenching

As discussed previously, QS systems play important roles in virulence factor production, biofilm formation, and motility. Therefore, how to inhibit QS is critical in combatting pathogenic bacteria and could provide an alternative for antibiotic treatment. The inhibition of QS is commonly called quorum quenching (QQ). In general, there are three strategies that can be used to inhibit QS: 1). inhibition of signal synthesis, 2). enzymatic degradation and inactivation of signaling molecules, 3). blocking signal reception (**Figure 1.8**).

Inhibition of signal synthesis can limit QS signal accumulation. In general, AHL signals are generated by acyl-HSL synthases, which are formed from the substrates S-adenosyl-methionine (SAM) and acyl-acyl-carrier-protein (acyl-ACP).[132-134] Therefore, suppressions of SAM biosynthesis, acyl-ACPs generation, or inactivation of the synthase enzyme are potential ways to inhibit AHL signal production.[132] For example, 5′-methylthioadenosine/S-adenosylhomocysteine nucleosidase (MTAN; also known as Pfs)

plays an important role in biosynthesis of both AHL and AI-2 autoinducers.[135] Therefore, MTAN inhibitors, such as immucillin A and DADMe-ImmA derivative, can be used to suppress the autoinducer production.[136,137]

Many AHL degrading or modifying enzymes have been reported and these can be classified into four groups: i). lactonases, which are able to open the homoserine lactone ring.[138] ii). acylases, which can hydrolyze AHLs at the amide bond and release the fatty acid tail group and homoserine lactone head group.[139] iii). oxidoreductases, which can convert the 3-oxo-substituted AHL to the cognate 3-hydroxyl-substituted AHL.[140] iv). Cytochrome oxidases, which catalyze oxidation of the acyl chain.[141] It has been found that some bacteria can degrade their own autoinducers and inhibit the QS activities when group behaviors are assumed to be problematic for the population in the late growth stage.

The third strategy of QQ is blocking the signal reception, which is based on different QS signal analogues. The basic idea is to generate analogues of native signaling molecules. The native signaling molecules can be modified to keep the ligand-receptor interaction properties while generating nonproductive ligand-receptor complexes that do not cause the same effect in gene expression. The competitive binding of the native signaling molecules is a good alternative for QQ. Various AHL analogues have been tested, such as lactams,[142] thiolactones,[143] urea,[144] and triazolyldihydrofuranone based analogues.[145]

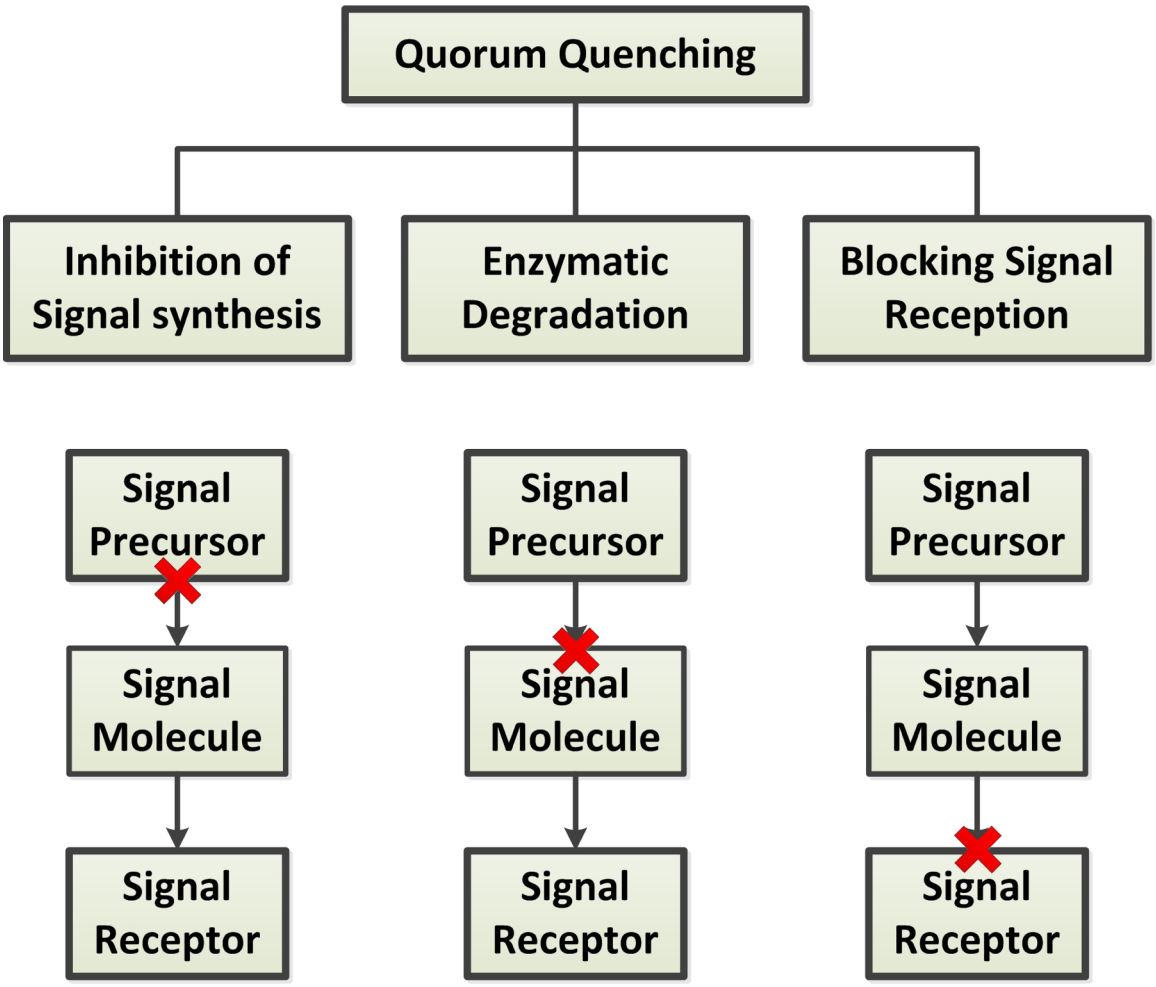

Figure 1.8 Quorum quenching strategies. The three different kinds of strategies are employed to inhibit signal synthesis, degrade or inactivate signaling molecules, and block signal reception, respectively.

1.4.2 Acyl-homoserine lactone lactonase and acylase

Enzymatic degradation of AHL signaling molecules is a promising strategy to inhibit pathogenic bacterial infections. AHL-lactonase and AHL-acylase are two major classes of enzymes that degrade QS molecules. As shown in **Figure 1.9** AHL-lactonase can target the homoserine lactone ring group of AHLs by hydrolyzing the ester bond to yield the corresponding acyl homoserine molecule. Such hydrolysis can occur spontaneously at alkaline pH and is reversible at acidic conditions. On the other hand, AHL-acylase can

break down the amide bond between the acyl chain and homoserine lactone head group of AHL molecules, which releases a homoserine lactone and a corresponding fatty acid.[139] In this study, we focus on the binding specificity of AHL acylase since it is difficult for the amide bond to be rebuilt after hydrolysis by AHL acylase. AHL-acylases are substrate specific enzymes, which prefer different AHL ligands based on the acyl chain length[146]. For example, AHL-acylase PvdQ enzyme from *P. aeruginosa* was studied in this research, which is much more sufficient to hydrolyze AHLs with acyl-chains longer than 10 carbon atoms.[146,147]

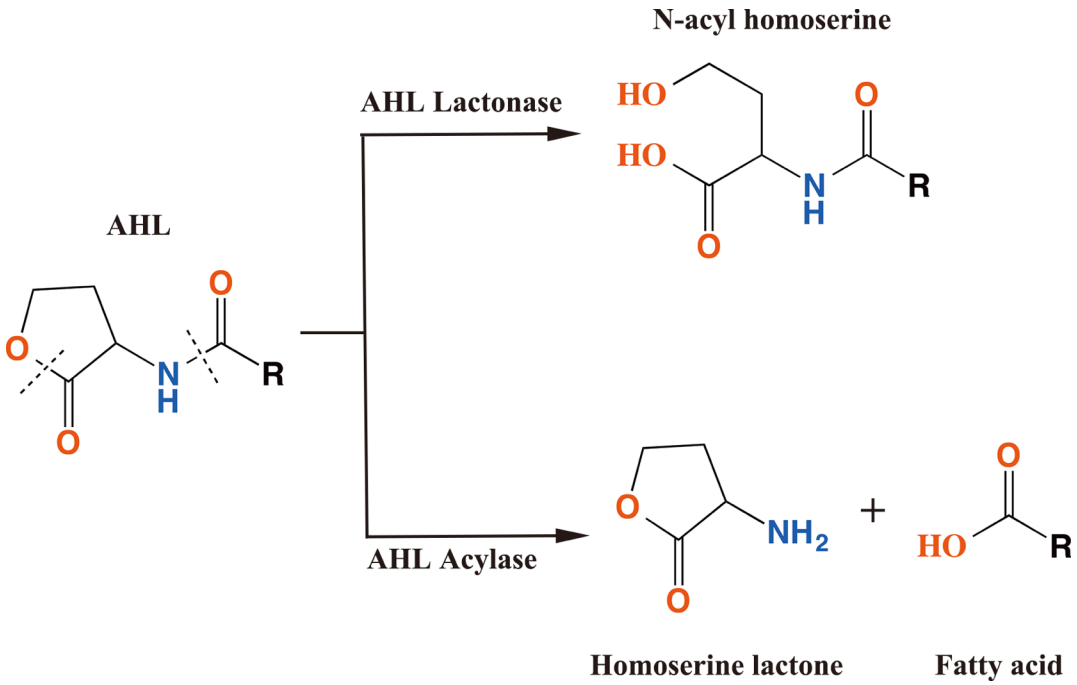

Figure 1.9 AHL signal degradation with lactonases and acylases. Dashed lines indicate the sites of bond cleavage. (Adapted from LaSarre B.[132])

1.4.3 Structure and function of PvdQ from *P. aeruginosa*

As discussed above, *P. aeruginosa* and *B. cenocepacia* are two of the most problematic pathogens associated with various serious infections, such as cystic fibrosis. Moreover,

the co-infection of these two pathogens makes the colonization even more difficult to eradicate. Since both *P. aeruginosa* and *B. cenocepacia* employ QS to control their actions, the targeting of QS is an attractive way to inhibit virulence and biofilm formation. 3-oxo-C12-HSL and C4-HSL are the two major endogenously generated communication molecules in *P. aeruginosa* that mediate the QS. C8-HSL is the signal molecule for *B. cenocepacia*. Therefore, degradation of AHL signaling molecules is a new way to inhibit bacterial infections. As mentioned earlier, AHL acylase PvdQ is one of the structurally well-characterized QQ enzymes, which can break down the amide bond between the AHL head group and the acyl tail group. PvdQ from *P. aeruginosa* is a typical Ntn-hydrolase with two chains interwoven together (**Figure 1.10**). Heterodimer PvdQ forms two protruding knobs (A-knob and B-knob), which forms a heart-shaped conformation with a solvent-accessible cleft in the center. PvdQ has 710 resides in total (missing 7 residues) and an unusually hydrophobic binding pocket.

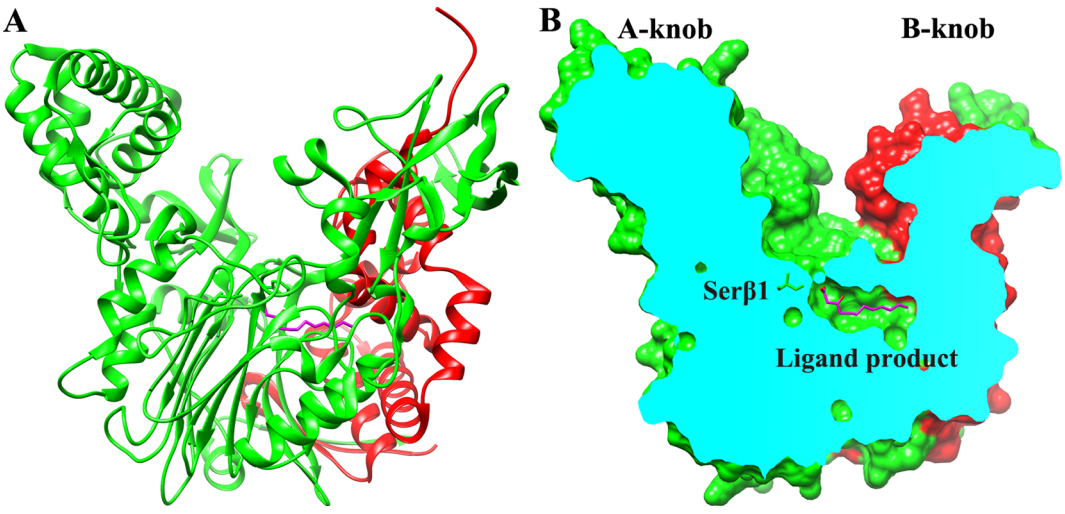

Figure 1.10 A: The 3D structure of PvdQ enzyme B: Sliced surface of PvdQ with substrate binding pocket. A chain colored in red, while B chain colored in green. The ligand product after hydrolyzed by PvdQ colored in magenta, while catalytic nucleophile Serβ1 colored in green.

PvdQ has two functions: siderophore biosynthesis and quorum quenching. Siderophores have high affinity for iron to scavenge free iron from surrounding environment. *P. aeruginosa* can produce different siderophores to acquire iron under iron limited environment.[83] PvdQ plays a key role in siderophore biosynthesis in order to maintain iron homeostasis. PvdQ is able to hydrolyze an *N*-myristic acid substituent from a precursor of the iron-scavenging pyoverdine, the major siderophore of *P. aeruginosa* (**Figure 1.11**).[148,149] Inhibition of PvdQ will block the iron acquisition and impair bacterial growth in an iron limited environment. Therefore, the study of the substrate binding specificity will pave the way for inhibitor design. As discussed above, the second function of PvdQ is acting as an AHL acylase with the ability to hydrolyze the amide bond of AHL signaling molecules used by various Gram-negative bacteria. PvdQ has an unusual hydrophobic binding pocket and exhibits different binding affinity toward different ligands. Four AHL substrates (3-oxo-C12-HSL and C4-HSL from *P. aeruginosa*, C8-HSL from *B. cenocepacia*, and C12-HSL as a control ligand) were investigated in this study in order to better understand the properties, structure, and mechanism of the PvdQ enzyme. Studying the performance of PvdQ on AHL substrates with different chain lengths and substituents can provide new insights and help to identify the factors related to the catalytic efficiency (**Figure 1.12**). The obtained information and analysis can be applied to target the bacterial infection through design of more an efficient and promiscuous PvdQ, new inhibitor against PvdQ, and other clinical or technological challenges.

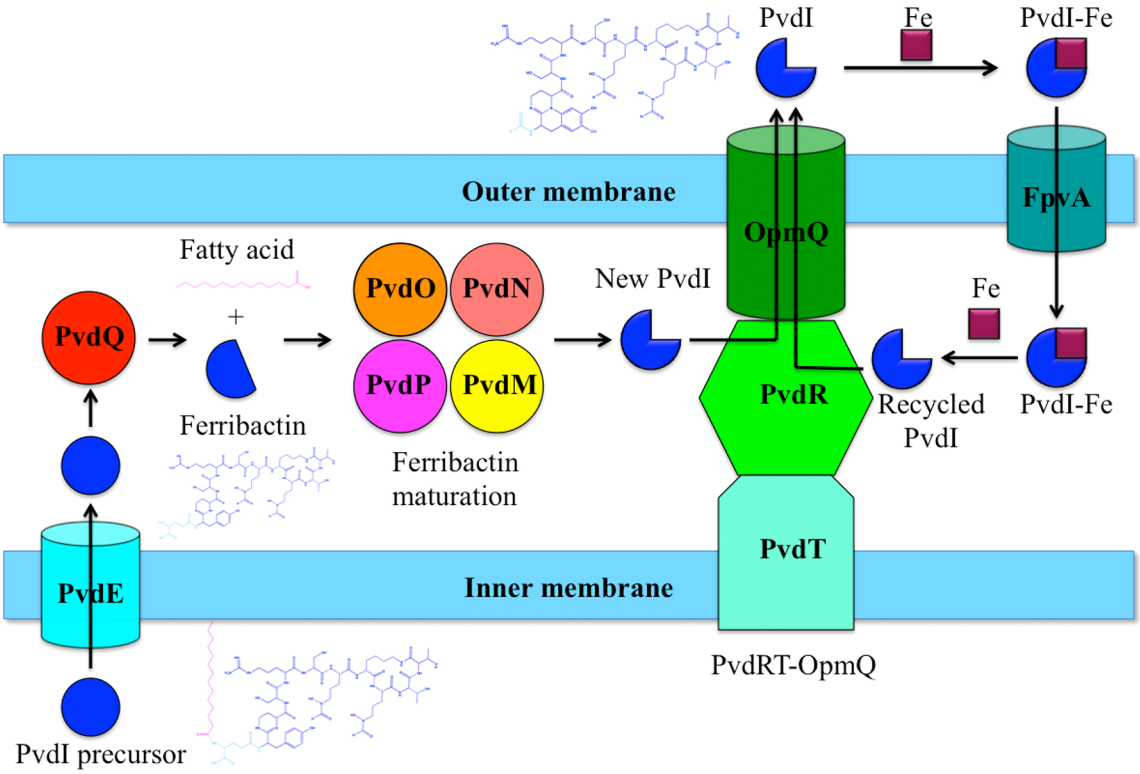

Figure 1.11 Siderophore biosynthetic function of PvdQ. PvdQ is able to hydrolyze an *N*-myristic acid substituent from a precursor of the iron-scavenging pyoverdine, PvdI precursor. After maturation, PvdI can scavenge free iron from surrounding environment and shuttle it back to the bacteria.

(Adapted from Nadal-Jimenez P.[148])

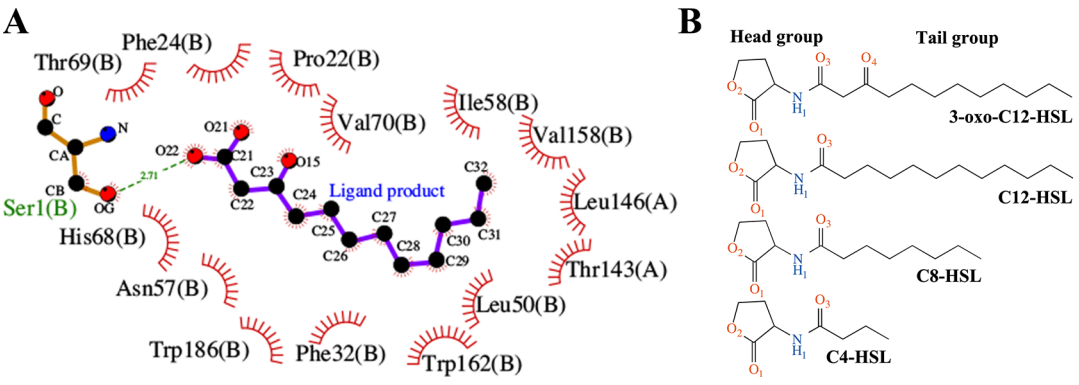

Figure 1.12 Substrate binding pocket and different ligands of PvdQ enzyme. A: The hydrophobic binding pocket of PvdQ. The hydrogen bonds are represented by dashed lines, while the hydrophobic interactions are indicated by arcs with spokes. B: Structures of four different ligands.

1.4.4 Mechanism of PvdQ enzyme

As a therapeutic enzyme, PvdQ plays an important role in interfering in quorum sensing, thereby inhibiting the virulence gene expression and biofilm formation. The heterodimer PvdQ is a typical Ntn-hydrolase (N-terminal nucleophile). It can activate autocatalytically and cleave an amide bond with a N-terminal catalytic nucleophile Serβ1. As represented in **Figure 1.13**, there are two proposed chemical mechanism of PvdQ. In path a., the α-amino group of Serβ1 is able to deprotonate its own hydroxyl group from Serβ1 directly to activate this catalytic center. In path b., water is involved in the activation of the Serβ1 nucleophile by relaying the proton from the Oγ to the alpha-amino group. After activation, the hydroxylate of the nucleophile Serβ1 can attack the carbonyl carbon of the scissile bond of the substrate. This leads to the formation of the first tetrahedral intermediate state, which is stabilized by the backbone of Valβ70 and side chain of Asnβ269. In path a. or path b., ammonia or water donates its proton to the amine of the HSL-leaving group, which results in the collapse of the tetrahedral intermediate state into the acyl-enzyme intermediate state. Then the α-amino group or water hydroxylate attack the carbonyl carbon of the acyl-enzyme intermediate state, which results in a second tetrahedral intermediate state stabilized by the backbone of Valβ70 and side chain of Asnβ269. The collapse of the second tetrahedral intermediate state results in the regeneration of the enzyme PvdQ.[148,149]

Lacking a functional PvdQ in *P. aeruginosa* will inhibit bacterial growth in an iron-limited environment, impair biofilm formation, and reduce virulence expression.[9] However, the details of the binding interactions of PvdQ with bacterial substrates are not fully understood.

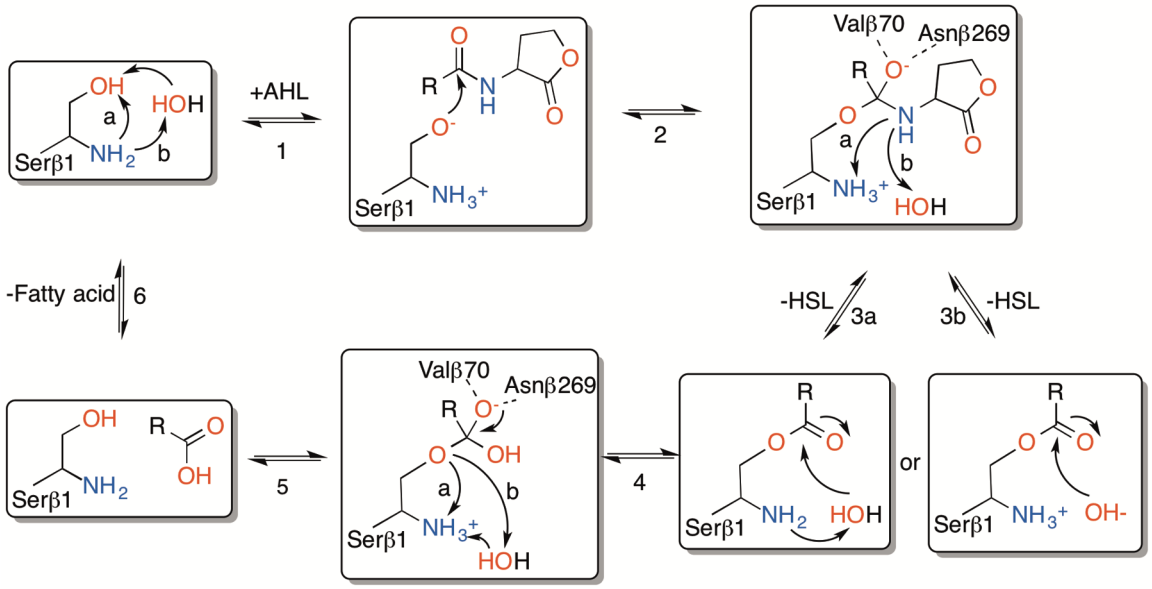

Figure 1.13 Chemical mechanism of PvdQ. There are two proposed chemical mechanisms of PvdQ. In path a., the α-amino group of Serβ1 is able to deprotonate its own hydroxyl group from Serβ1 directly to activate this catalytic center. In path b., water is involved in the activation of the Serβ1 nucleophile by relaying the proton from the Oγ to the alpha-amino group. Two tetrahedral intermediate states and one acyl-enzyme intermediate state are formed. The collapse of the second tetrahedral intermediate state results in the regeneration of the enzyme PvdQ. Serβ1 is the catalytic nucleophile. R is the acyl chain, which is released as a fatty acid.

(Adapted from Clevenger K. D.[207])

1.5 Hypothesis and scope of research

In this study, the prediction of the binding affinity and dynamic motions between the PvdQ enzyme and different bacterial substrates would provide a new insight into the determinants of substrate preference and substrate discrimination of PvdQ. Therefore, I propose that the mutations of some specific residues, such as Serβ1, Pheβ24, Leuα146, Asnβ57, will affect the accessibility and volume of the substrate binding pocket. And I also believe that the structural communication between the A/B knob and active site of

PvdQ complex will influence the orientation of the side chain of residues lining the binding pocket and thereby affecting the correct arrangement of the active site toward different substrates. This study will be helpful in the design of an efficient PvdQ enzyme with broader substrate promiscuity or in the design of inhibitors of PvdQ that can inhibit the iron acquisition.

In summary, the co-infection of the two pathogens *P. aeruginosa* and *B. cenocepacia* are difficult to eradicate and can be deadly for hospitalized patients. The PvdQ enzyme can be used to interfere the QS, thereby inhibiting these bacterial virulence factor production and biofilm formation. PvdQ has different preference for QS signaling molecules with different acyl chain lengths and substituents. The substrate binding specificity determinants of the quorum-quenching enzyme PvdQ with the different bacterial ligands are unknown and unintuitive. The starting point of this study is to provide a better understanding of the dynamical characteristics of the PvdQ complex; thus pave the way to improve the catalytic efficiency of PvdQ or the future design of new inhibitors. In this study, multiple molecular docking calculations are performed to study the favorable binding poses and binding affinities of PvdQ with different bacterial ligands. Various molecular dynamic simulation methods are implemented to further gain insight into the determinants of the substrate binding specificity of PvdQ, which can lead to mutants with higher efficiency and broader substrate promiscuity.

Chapter 2 Theoretical background and methods

2.1 pKa prediction

Protein stability and function are affected by the pH of the surrounding environment, which can exchange protons with the titratable residues of the protein to influence the ionization state of residues and the charge properties of the protein. In most biological reactions, the prediction of the correct ionization state of acidic or basic residues in a protein is critical to understand the function of the molecule. pKa values of ionizable residues can be used to measure proton affinity and strength of an acid for a given pH. In other words, pKa is the expression of the acid dissociation constant (K_a) on the logarithmic scale. The K_a is used to show the strength of the acid. A strong acid can dissociate completely in water, while a weak acid dissociates incompletely. An acid, HA, can dissociate into the conjugate base, A^- and proton, H^+. K_a is expressed by the product of the concentrations of $A^-$and H^+ divided by the concentration of HA.[150] pKa is an easy and convenient way to represent the K_a. The K_a of a weak acid HA is a constant value at a specific pH and enviroment. The relationship between the pKa and pH is shown in the following equations:

$$HA \rightleftharpoons A^- + H^+ \qquad \text{(Equation 2.1)}$$

$$K_a = \frac{[A^-][H^+]}{[HA]} \qquad \text{(Equation 2.2)}$$

$$pKa = -\log_{10} K_a \qquad \text{(Equation 2.3)}$$

$$pH = pKa + \log_{10} \frac{[A^-]}{[HA]} \qquad \text{(Equation 2.4)}$$

Therefore, pKa equals to pH when half of the acid is dissociated ($[A^-]/[HA]=1$). **Table 2.1** shows the pKa values of some common amino acids found in proteins when these residues are alone in water.[151] pKa plays an important role in protein catalysis, protein stability, and protein-ligand interaction. It can help to predict the accurate electrostatic potentials of the protein and understand the pH dependent protein recognition and enzyme catalysis. The pKa of amino acid is represented by the pKa of the side chain of the amino acid. However, the pKa of an isolated amino acid can be different than the pKa of the residue in a protein due to the effects of the surrounding environment. The main factors affecting pKa values in proteins are desolvation of the residue by nearby residues and the electrostatic interactions (or lack) with surrounding residues. Therefore, the prediction of pKa of a residue in a protein at a specified condition is extremely important.

Table 2.1 pKa's of common amino acids in water

	Name	pKa of α-COOH	pKa of α-$NH3^+$	pKa of side chain
Non-polar amino acids	Gly	2.35	9.78	-
	Ala	2.35	9.87	-
	Val	2.29	9.74	-
	Leu	2.33	9.74	-
	Ile	2.32	9.76	-
	Met	2.13	9.28	-
	Phe	2.20	9.31	-
	Pro	1.95	10.64	-
	Trp	2.46	9.41	-
Polar amino acids	Asn	2.14	8.72	-
	Gln	2.17	9.13	-
	Ser	2.19	9.21	13
	Thr	2.09	9.10	13
	Cys	1.92	10.70	8.37
	Tyr	2.20	9.21	10.46
Basic amino acids	Lys	2.16	9.06	10.54
	Arg	1.82	8.99	12.48
	His	1.80	9.33	6.04
Acidic amino acids	Asp	1.99	9.90	3.90
	Glu	2.10	9.47	4.07

In this study, the pKa was predicted via the University of Houston Brownian Dynamics (UHBD) program.[152] UHBD is a program capable of providing fast evaluations of electrostatic potential values for all ionizable residues in a protein as well as the pH titration behaviors of the titratable residues. The finite-difference Poisson-Boltzmann method (FDPB) method is used in UHBD to numerically solve the linearized Poisson-

Boltzmann equation:[153]

$$\nabla \cdot \varepsilon(r)\nabla\phi(r) = -4\pi\rho^f(r) + \lambda(r)\bar{k}^2\phi(r) \qquad \text{(Equation 2.5)}$$

where ε is the dielectric constant function, ϕ is the electrostatic potential at position r to the density of the charge distribution ρ^f at position r, while λ indicates the accessible regions to ions and $\bar{k}$ is a modified Debye-Hückel parameter.

In UHBD, the predictions of the ionization states are obtained from the difference in the electrostatic work of changing the ionizable site from the neutral to the ionized state in the protein and the electrostatic work of making the same alteration in the model compound.

For a given pH, the free energy change for protonation of a single titratable group can be obtained from

$$\Delta G = 2.303RT(\text{pH} - \text{pKa}) \qquad \text{(Equation 2.6)}$$

It should be noted that $\text{pK}_{\text{a,model}}$ should be used if the group is embedded in the model amino acid and $\text{pK}_{\text{a,intrinsic}}$ should be used if the group is protonated in an otherwise neutral protein. $\text{pK}_{\text{a,intrinsic}}$ is represented as:

$$\text{pK}_{\text{a,intrinsic}} = \text{pK}_{\text{a,model}} - \gamma\Delta\Delta G/2.303RT \qquad \text{(Equation 2.7)}$$

where

$$\gamma = \begin{cases} -1, & \text{for an acidic site} \\ +1, & \text{for a basic site} \end{cases} \qquad \text{(Equation 2.8)}$$

and

$$\Delta\Delta G = \Delta G^{(el)}_{protein} - \Delta G^{(el)}_{model} \qquad \text{(Equation 2.9)}$$

$\Delta G^{(el)}_{protein}$ and $\Delta G^{(el)}_{model}$ represent the free energy differences for ionization of a given site in the model amino acid and in the protein with all other residues neutral, respectively. In this study, predicting pKa with UHBD is important to better understand the stability and catalytic mechanisms of a protein.

2.2 Molecular docking calculations

2.2.1 Outline and principle

Molecular docking is a method can be used to predict the best matching between two molecules. Molecular docking calculations can be applied to study the protein-substrate interactions, binding-site identification, catalytic mechanisms, virtual screening, and drug discovery. For example, protein-ligand docking can be used to predict the possible binding modes and binding affinity to better understand the biological activity of the protein-ligand complex. In general, the molecular docking procedure includes two components: a search algorithm and a scoring function. The search algorithm finds different structural conformations for the ligand by using methods such as: Monte Carlo methods, genetic algorithms, exhaustive search methods, fragment-based methods, tabu searches, and systematic searches, etc.[154] It is difficult to decide whether the orientations of the generated ligand poses fit the active site of the receptor. Therefore, a scoring procedure is needed to measure the fitness of a substrate into the binding site. The

primarily used scoring methods include empirical, knowledge based, and force-field based scoring functions. Many docking programs are currently available, such as AutoDock,[155] AutoDock Vina (Vina),[156] Schrödinger Glide,[157] GOLD,[158] LeDock,[159] FlexX,[160] UCSF Dock,[161] LigandFit,[162] MOE Dock,[163] and Surflex.[164] The differences between those methods are based on their different search and/or scoring functions. In practice, it is generally necessary to combine different docking methods to achieve better predictions.[165] In this study, three of the most popular methods (i.e., Schrödinger induced fit docking (IFD), GOLD, and AutoDock Vina) are employed to obtain a more reliable cross-validated docking result. IFD is based on Glide (exhaustive search based) and the Refinement module in Prime.[166] GOLD is a genetic algorithm based docking program with flexible ligand and partial flexibility for the protein. Autodock Vina uses a Carlo sampling method and the Broyden-Fletcher-Goldfarb-Shanno (BFGS) approach for local optimization.[167]

2.2.2 Search algorithms

The difficulty of the searching procedure in docking is in part due to the fact that many degrees of freedom are involved: six degrees of freedom of the translation and rotation of one molecule relative to the other; additional conformational degrees of freedom for both protein and ligand; the solvent can also affect the protein-ligand geometry. Therefore, the search algorithm needs to generate extensive poses of the molecule in the binding site.

2.2.2.1 Exhaustive search

The exhaustive search method is used in docking by rotating all potential rotatable bonds of the ligand in the binding site at a specified interval. The computational burden of

exhaustive search can become too heavy as the search space becomes very large. In Glide, heuristics are used to focus on the conformational space that tends to have favorable ligand poses and a grid representation of target's shape and properties is generated as well as an initial set of low-energy ligand conformations in ligand torsion-angle space.[168] Moreover, an initial screening of favorable ligand poses is carried out to further reduce the conformational space using approximate positioning and scoring methods.[157,168] Then a high-resolution docking search is implemented, which minimizes the molecular mechanics energy function; then uses a Monte Carlo algorithm to further examine the nearby minim.[168]

2.2.2.2 Genetic algorithm

The general idea of a genetic algorithm is based on Darwin's theory of evolution and natural selection. The set of randomly chosen ligand placements in the conformational space can be seen as the 'populations'. Degrees of freedom of the ligand are represented as 'genes', which make up the 'chromosomes' that represent the ligand's poses.[169] Genes describe the ligand placement at a certain position around the protein with variables such as: translation, rotation, and ligand torsions. Genetic operators, such as mutation, crossover, and selection are also used in genetic algorithm. Mutation changes the genes randomly and crossover exchanges genes between two chromosomes.[169] After the 'genetic operator', a new ligand structure is obtained and can be assessed by scoring functions (i.e., selection operator). The selected structures are then used for the next generation and the procedure run iteratively until some stopping criteria (i.e., number of runs, or energy difference between two consecutive runs) are met.

2.2.2.3 Monte Carlo method

Monte Carlo methods produce an initial pose of the ligand in the binding site with random conformation, bond rotation, rigid-body translation or rotation.[170,171] The obtained conformation is scored with some energy-based selection criteria. If it passes the criteria, it will be saved and some changes will be made to generate a new conformation, which will again be scored according to the same criteria. If the new conformation gets a better score it is saved otherwise it is accepted or rejected via a Metropolis test.[168] If the new conformation is not a new minimum, a Boltzmann-based probability function is implemented.[168] If it passes the probability test, the conformation is accepted; otherwise, it is rejected. The iterations continue until the defined number of conformations is generated.[169]

2.2.3 Scoring algorithms

The scoring algorithms in docking can generally be divided in three categories: force-field based, empirical, and knowledge based scoring functions. Different docking programs have different scoring strategies. For example, Glide uses empirical scoring, GOLD uses force-field based scoring, and AutoDock Vina uses knowledge based scoring. Many docking programs also combine different scoring functions to improve the performance. For example, AutoDock Vina combines both knowledge-based and an empirical approach. In this section, examples of scoring algorithm are introduced for each category.

2.2.3.1 Empirical scoring algorithm

The empirical scoring function includes several energy terms that are based on

experimentally observed values. The ChemScore function[157,172] used in Glide is an empirical based scoring algorithm with form of

$$\Delta G_{\text{bind}} = C_0 + C_{\text{lipo}} \sum f(r_{lr}) + C_{\text{hbond}} \sum g(\Delta r) h(\Delta \alpha) + C_{\text{metal}} \sum f(r_{lm}) + C_{\text{rotb}} H_{\text{rotb}}$$

(Equation 2.10)

Each component on the right represents the descriptions of specific interactions such as lipophilic interaction, hydrogen bonding interactions, interaction with metal, and others. Based on ChemScore, two more complicated GlideScore algorithms (i.e., GlideScore Standard-Precision (SP) and GlideScore Extra Precision (XP)) are also used in Glide.[157] GlideScore SP[157] is a more forgiving function and extends the ChemScore function as

$$\Delta G_{\text{bind}} = C_{\text{lipo}} \sum f(r_{lr}) + C_{\text{hbond-neut-neut}} \sum g(\Delta r) h(\Delta \alpha) +$$

$$C_{\text{hbond-neut-charged}} \sum g(\Delta r) h(\Delta \alpha) +$$

$$C_{\text{hbond-charged-charged}} \sum g(\Delta r) h(\Delta \alpha) +$$

$$C_{\text{max-metal-ion}} \sum f(r_{lm}) + C_{\text{rotb}} H_{\text{rotb}} +$$

$$C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} +$$

$$C_{\text{vdW}} E_{\text{vdW}} + \text{solvation terms} \quad \text{(Equation 2.11)}$$

The first lipophilic term is same as ChemScore. The hydrogen-bonding term is divided into three components based on whether the donor and acceptor are neutral or charged. The fifth term represents the mental-ligand interaction term. The seventh term rewards instances in which a polar but non-hydrogen-bonding atom is found in a hydrophobic region. The other major components are the contributions from the Coulomb and van der Waals (vdW) interaction energies. The remaining major component is related to the solvation model.[157]

2.2.3.2 Force-field based scoring algorithm

The force field describing the energy of the system can be divided into different terms: non-bonded interaction terms (i.e., van der Waals and electrostatic interactions) and bond stretching/bending/torsional energy.[168] Force-field based scoring algorithms use a variety of force-field parameters. The Goldscore [173] scoring function used in GOLD has the form of

$$\text{Goldscore} = S_{hb\_ext} + S_{vdw\_ext} + S_{hb\_int} + S_{vdw\_int} \qquad \text{(Equation 2.12)}$$

where S_{hb_ext} represents the protein-ligand hydrogen bond score, S_{vdw_ext} is the van der Waals score, S_{hb_int} is the intramolecular hydrogen bonds in the ligand. S_{vdw_int} is the contribution to the fitness caused by intramolecular strain in the ligand. GOLD can place the ligand in the binding site by adding fitting points and it is optimized for the prediction of binding positions.[174]

2.2.3.3 Knowledge based scoring algorithm

The scoring function used in AutoDock Vina is introduced in this section. The conformation-dependent part of the scoring function has the form of[156]

$$c = \sum_{i<j} f_{t_i t_j}(r_{ij}) \quad \text{(Equation 2.13)}$$

where t_i represents the type of atom i, r_{ij} is the interatomic distance between atom i and atom j. $f_{t_i t_j}$ is the interacting function with form of

$$f_{t_i t_j}(r_{ij}) \equiv h_{t_i t_j}(d_{ij}) \quad \text{(Equation 2.14)}$$

where $d_{ij} = r_{ij} - R_{t_i} - R_{t_j}$ is the surface distance.[174] R_t is the van der Waals radius of atom type t. $h_{t_i t_j}$ represents a weighted summation of steric interactions, hydrophobic interaction, and hydrogen bonding. The steric interactions are same for all atom pairs and contain three components:[156]

$$\text{gauss}_1(d) = e^{-(d/0.5\text{Å})^2} \quad \text{(Equation 2.15)}$$

$$\text{gauss}_2(d) = e^{-(\frac{d-3\text{Å}}{2\text{Å}})^2} \quad \text{(Equation 2.16)}$$

$$\text{repulsion}(d) = \begin{cases} d^2, & \text{if d} < 0 \\ 0, & \text{if } d \geq 0 \end{cases} \quad \text{(Equation 2.17)}$$

The hydrophobic term and hydrogen bonding term are both linear functions in different distance segments. The scoring value in Equation 2.13 can also be divided into the sum of intermolecular and intramolecular contributions as:

$$c = c_{inter} + c_{intra} \quad \text{(Equation 2.18)}$$

An optimization algorithm is used in Vina to find the global minimum of c and other

low-scoring conformations. The predicted binding free energy can be determined from the intermolecular part of the lowest-scoring pose as[156]

$$s_1 = g(c_1 - c_{intra1}) = g(c_{inter1}) = \frac{c_{inter1}}{1+wN_{rot}} \qquad \text{(Equation 2.19)}$$

where g is a conformation-independent function with an increasing smooth nonlinear form. N_{rot} is the number of active rotatable bonds between heavy atoms in the ligand and w is the associated weight.

In practice, Vina ranks the conformations according to Equation 2.18. The scoring function in Vina combines both knowledge-based potentials and empirical scoring functions: both the conformational preferences of the receptor-ligand complexes and the experimental affinity measurements provide empirical information.

2.2.4 Docking protocols

In this section, the protocols of three docking programs used in this study are introduced.

2.2.4.1 Schrodinger induced fit docking (IFD)

It is known that the conformation of the binding site of a protein depends heavily upon structural changes caused by the binding of a ligand. Schrodinger's IFD protocol deals with this problem by combining Glide and Prime to process possible binding modes and the corresponding conformational changes.[175] First, IFD implements Glide to dock the ligand using a softened potential and removal of side chains. The van der Waals radii and an increased Coulomb-vdW cutoff are used in the procedure to temporarily remove highly flexible side chains. The potential is softened using information from side-chain flexibility. Many poses are returned from several docking runs and are clustered to obtain

representative poses. Then a Prime side-chain prediction is performed on residues within a specific distance to accommodate the ligand by reorienting nearby side-chains. A minimization procedure is carried out to these residues and the ligand. Finally, implementing Glide to re-dock each protein-ligand complex into its associated low energy structures and the resulting complexes are ranked according to GlideScore. The default Glide settings is used to rigorously dock the ligand into the induced-fit binding site. The output poses are scored with a new scoring function.[175] Therefore, IFD can combine the advantages of Glide's sophisticated scoring function and Prime's advanced conformational refinement.

2.2.4.2 GOLD docking

Similar with other docking programs, GOLD docking procedure also contains three components: a search algorithm, a scoring function, and a mechanism for placing the ligand in the binding site.[173] As introduced in Section 2.2.2.2, GOLD uses a genetic algorithm to explore possible binding modes. In the genetic algorithm, the following parameters are optimized: (a). dihedrals of ligand rotatable bonds; (b). dihedrals of protein OH groups and NH_3^+ groups; (c). ligand ring geometries by flipping ring corners; and (d). the mappings of the fitting points.[173] GOLD uses a unique fitting points-based method to place the ligand in the binding site. The fitting points are added to hydrogen bonding groups on protein and ligand, and acceptor points are mapped on the ligand on donor points in the protein and vice versa. Moreover, hydrophobic fitting points in the protein cavity are also created by GOLD where ligand CH groups are mapped.[173] The scoring function used in GOLD to rank different binding modes is a molecular mechanics-like function as introduced in Equation 2.12.

2.2.4.3 AutoDock Vina docking

Vina is designed to be compatible with AutoDock software to take advantage of the existing AutoDock Tools. Vina uses a model that predicts noncovalent bonds, hydrophilic, and hydrophobic interactions between the protein and ligand. The scoring function introduced in section 2.2.3.3 is used to calculate the free energy of the system. Vina places the ligand in a set conformation within the research region and estimates the binding affinity using the distances between the atoms of interest in the protein-ligand complex.[176] The search area needs to be specified by size and coordinates in the 3D space. If the search area is not big enough, the ligand might not be able to rotate properly. If the search area is too big the processing time becomes very long. AutoDock Tools can be used to define the search area. Then the ligand is positioned in a new different conformation in the search area, and the binding affinity is recalculated.[176] This procedure is repeated for all the conformations, which are scored based on the binding affinity. It should be noted that the outputs of Vina can be different for different runs since a random seed is used to decide which configurations are used, which means sometimes it is necessary to repeat Vina docking procedures to obtain converged results.

2.3 Molecular dynamics (MD) simulations

2.3.1 Outline and principle

The basic idea of MD simulation is to mimic the motions of atoms using computational tools. In general, two basic physic theories, such as Quantum Mechanics (QM) and Classical Mechanics (CM), can be used to compute the molecular dynamics. QM can accurately model the movement of electrons in a molecule by solving the wave functions.

However, QM is computationally expensive and is generally only suitable for small systems (< 100 atoms). Compared with QM, MD simulation with CM is less accurate. But the simulation is much faster and adaptive to large systems (>104 atoms) with long time duration (> 10 ns). Therefore, in this study, MD simulation with CM is used. MD simulation is a useful method to better understand the molecular basis of protein dynamics, flexibility, function, and stability.

For a set of atoms, the force experienced by any atom can be calculated from the energy function if the positions of the other atoms are also known. How the forces affect the atoms' motions can be determined by Newton's laws. In practice, continuous time can be divided into small discrete steps. At each step, the forces experienced by each atom can be calculated from the molecular mechanics force field. The position and velocity of each atom can be updated by Newton's laws of motion. Then the new positions and velocities can be used for the next step to repeat the procedure. Finally, the molecular dynamics motions can be simulated.

2.3.2 Potential energy functions

In order to estimate the forces between particles, a potential energy function is needed, which exists between bonded neighbors and non-bonded atoms. For bonded atoms, three types of interactions exist, i.e., stretching along the bonds, bending between bonds, and rotating around bonds. The bonded potentials include bond-length, bond-angle, proper dihedral, and improper dihedral potentials. In additional to the van der Waals and electrostatic potentials, the schematic representations for all the six sources of energy are shown in **Figure 2.1**.[177]

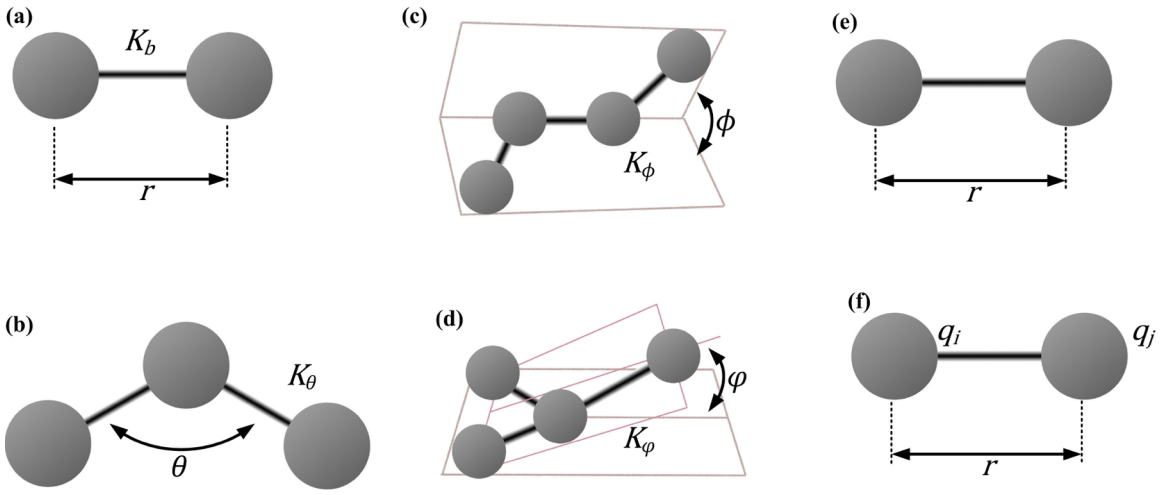

Figure 2.1 Different potentials. (a) bond length potential. (b) bond angle potential. (c) proper dihedral potential. (d) improper dihedral potential. (e) Van der Waals. (f) Electrostatic interactions. (Adapted from Maksim K.[177])

The bond-length potential controls the length of covalent bonds. The most common to be used is the harmonic bond potential, which is given by

$$V_{bond} = K_b(r - r_0)^2 \qquad \text{(Equation 2.20)}$$

where K_b is the force constant, r is the distance between covalently bonded atoms, r_0 is the equilibrium bond length.

As shown in **Figure 2.1 (b)**, the angle potential is defined by two bonds that share a common atom. Two most commonly used angle potentials are the harmonic and the cosine harmonic potential functions, which can be expressed as

$$V_{angle} = K_\theta^H(\theta - \theta_0)^2 \qquad \text{(Equation 2.21)}$$

and

$$V_{angle} = K_{\theta}^{CH}(\cos\theta - \cos\theta_0)^2 \qquad \text{(Equation 2.22)}$$

where θ is the bond angle, θ_0 is the reference angle, K_{θ}^{H} and K_{θ}^{CH} are the force constants for the harmonic and cosine harmonic functional forms, respectively. Both the force constants have a different order of magnitude and different units. In practice, Chemistry at HARvard Molecular Mechanics (CHARMM) uses the harmonic potential function and GROMOS96 uses the cosine harmonic potential functions.

Torsion potentials contains two cases, i.e., proper dihedral potential and improper dihedral potential as shown in **Figure 2.1 (c)** and **(d)**. Both potentials depend on a structure of four atoms, bonded in different ways. The proper dihedral angle potential contains four atoms bonded consecutively while the improper dihedral has structure of three atoms centered around a fourth atom. The major difference between these two potentials is how to define the torsional angles and the forms of the potential functions. As shown in **Figure 2.1 (c)**, the torsion angle for proper dihedral is defined by the angle of the two planes of the dihedral structure. For improper dihedral potential the torsional angle is defined by the angle between planes as shown in **Figure 2.1 (d)**. The cosine form of the dihedral potential can be expressed as

$$V_{proper\ dihedral} = K_{\phi}^{C}[1 + cos(n\phi - \phi_0)] \qquad \text{(Equation 2.23)}$$

where K_{ϕ}^{C} is the force constant, n is the multiplicity, which is a positive nonzero integer number describes the number of minima as the bond is rotated through 360^{o}. ϕ is the torsional angle and ϕ_0 is the angle where the potential passes the minimum value. The functional form of the improper dihedral can be expressed as

$$V_{improper\ dihedral} = K_{\varphi}(\varphi - \varphi_0)^2 \qquad \text{(Equation 2.24)}$$

where K_{φ} is the force constant, φ is the torsional angle and φ_0 is the equilibrium value. The dihedral potential is mainly used to describe the rotation around a bond and the improper dihedral potential is mostly used to maintain planarity in a molecular structure.

For non-bonded atoms, two potential functions need to be included, i.e., van der Waals potential and electrostatic potential. The most commonly used approximation for the van der Waals potential is the Lennard-Jones (LJ) potential, which can be expressed by

$$V_{LJ}(r) = 4\varepsilon\left[(\frac{\sigma}{r})^{12} - (\frac{\sigma}{r})^{6}\right] \qquad \text{(Equation 2.25)}$$

where ε is the depth of the potential well, σ is the finite distance at which the potential is zero, r is the distance between the particles. The LJ potential versus the distance between two particles is shown in **Figure 2.2**. Particles with no net electrostatic charge will still tend to attract each other at short distance given their distance is not too close. Once the distance is too close, the atoms have overlapping electron clouds and start repel each other with strong force.

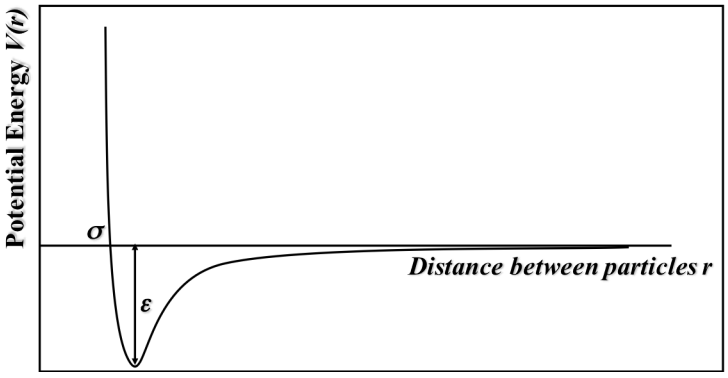

Figure 2.2 Lennard-Jones potential energy versus distance between atoms.

The electrostatic potential can be expressed with Coulomb's law as

$$V_{Elec} = \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r} \qquad \text{(Equation 2.26)}$$

where ε_0 is the permittivity of free space, ε is the dielectric constant, q_i and q_j are the signed magnitudes of the charges, and r is the distance between the charges. If the charges have opposite signs, the force of the interaction is attractive and vice versa.

In summary, the total potential energy is the summation of all the above six categories of potentials and can be given by[178]

$$V = \sum_{bonds} K_b(r - r_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} K_\phi[1 + cos(n\phi - \phi_0)] + \sum_{impropers} K_\varphi(\varphi - \varphi_0)^2 + \sum_{i>j} \varepsilon \left[\left(\frac{r_m}{r}\right)^{12} - 2\left(\frac{r_m}{r}\right)^{6}\right] + \sum_{i>j} \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r} \qquad \text{(Equation 2.27)}$$

2.3.3 Numerical integration

The previous section introduced the forms of the potential energies, which is a function of 3N atomic positions and determines the forces as

$$f_i = -\frac{\partial V}{\partial r_i} \qquad \text{(Equation 2.28)}$$

where f_i represents the forces acting on the i-th atom and is derived from potential energy $E(r^N)$ and $r^N = (r_1, r_2, \ldots, r_N)$ represents the 3N atomic coordinates. N is the number of atoms. By knowing the force, the molecular dynamics simulations can be achieved based on Newton's law of motion: $F = ma$, where F is the force on an atom, m is the mass of the atom, and a is the acceleration. It is known that acceleration is the derivative of

velocity v, which is the derivative of position r such as

$$\begin{cases} v = \frac{dr}{dt} \\ a = \frac{dv}{dt} = \frac{F}{m} \end{cases} \qquad \text{(Equation 2.29)}$$

Combining Equation 2.28 and 2.29 gives the relation of potential energy and motions. It should be noted that the potential energy is a function of the positions of all the atoms in the system and there is no analytical solution due to the high complexity. In practice, numerical methods have been used to perform the integrations. For a small time step δt, Taylor expansions of the position function r(t) can be written as

$$r(t+\delta t) = r(t) + \frac{dr(t)}{dt}\delta t + \frac{d^2r(t)}{dt^2}\frac{\delta t^2}{2} + \cdots = r(t) + v(t)\delta t + \frac{f(t)}{m}\frac{\delta t^2}{2} + \cdots$$

(Equation 2.30)

and

$$r(t-\delta t) = r(t) - \frac{dr(t)}{dt}\delta t + \frac{d^2r(t)}{dt^2}\frac{\delta t^2}{2} + \cdots = r(t) - v(t)\delta t + \frac{f(t)}{m}\frac{\delta t^2}{2} + \cdots$$

(Equation 2.31)

Combining Equation 2.30 and 2.31 we have

$$r(t+\delta t) \approx 2r(t) - r(t-\delta t) + \frac{f(t)}{m}\delta t^2 \qquad \text{(Equation 2.32)}$$

Each position is obtained from the current and previous positions. Equation 2.32 is

known as the Verlet algorithm[179]. The Verlet algorithm is straightforward and needs modest storage. The disadvantage is lacking of explicit velocity term. A reformed Verlet algorithm that uses the velocity directly is known as velocity Verlet algorithm with expressions as

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{f(t)}{2m}\delta t^2 \qquad \text{(Equation 2.33)}$$

$$v(t+\delta t) = v(t) + \frac{f(t+\delta t)+f(t)}{2m}\delta t \qquad \text{(Equation 2.34)}$$

The basic steps of the Verlet algorithm as the following:

1). The forces on every atom can be calculated using the force field given the position $r(t)$ and velocity $v(t)$ at time t.

2). Using Equation (18) to get an update of the position $r(t+\delta t)$.

3). Calculate the new force $f(t+\delta t)$ using the updated position $r(t+\delta t)$.

4). Using Equation (19) to get an update of the velocity $v(t+\delta t)$.

5). Go back to step 1.

Another modified Verlet algorithm is known as the leap-frog algorithm[180] which solves for the velocities at half time step intervals:

$$v(t+\delta t/2) = v(t-\delta t/2) + \frac{f(t)}{m}\delta t \qquad \text{(Equation 2.35)}$$

$$r(t+\delta t) = r(t) + v(t+\delta t/2)\delta t \qquad \text{(Equation 2.36)}$$

The advantage of the leap-frog algorithm is that the velocity is calculated explicitly. One

disadvantage of the leap-frog algorithm is that the velocity is not calculate at the same time as position, making it difficult to calculate the kinetic energy contribution to the total energy at any one point in time.

One important configuration of the integration algorithm is the proper selection of the time step δt. Too large a time step causes the simulation become unstable since a large time step may make the positions of two atoms to be too close to each other and lead to strong forces. Too small a time step causes the trajectory covers only a limited part of the phase space. In general, the time step should be around one order of magnitude smaller than the fastest time scale in the system. However, there is no hard rule. In practice, **Table 2.2**[181] can be used for reference to select the time steps in the MD simulation.

Table 2.2 Timescales of molecular motions in MD simulations

Motions (distances, time)	Timescales (s)
Local motions (0.01 – 5 Å, 10^{-15} – 10^{-1} s) Atomic fluctuations Side chain motions Loop motions	10^{-15} - 10^{-12}
Rigid body motions (0.01 – 10 Å, 10^{-9} – 1 s) Helix motions Subunit motions Domain motions	10^{-12} - 10^{-9}
Medium-scale motions (> 5 Å, microseconds) Helix coil transitions Dissociation/association	10^{-9} - 10^{-6}
Large scale (> 5 Å, 10^{-7} – 10^{4} s) Protein folding and unfolding Protein interactions	10^{-6} - 10^{-1}

2.3.4 Simulation setup

A flowchart of the basic MD simulation is shown in **Figure 2.3**. The basic MD

simulation steps include initial molecule coordinates and velocities, structure minimization, solvation modeling, raising system temperature, equilibration to ensure system is stable, dynamics simulations under desired conditions (NVE, NPT, etc.), and final analysis to evaluate the results.

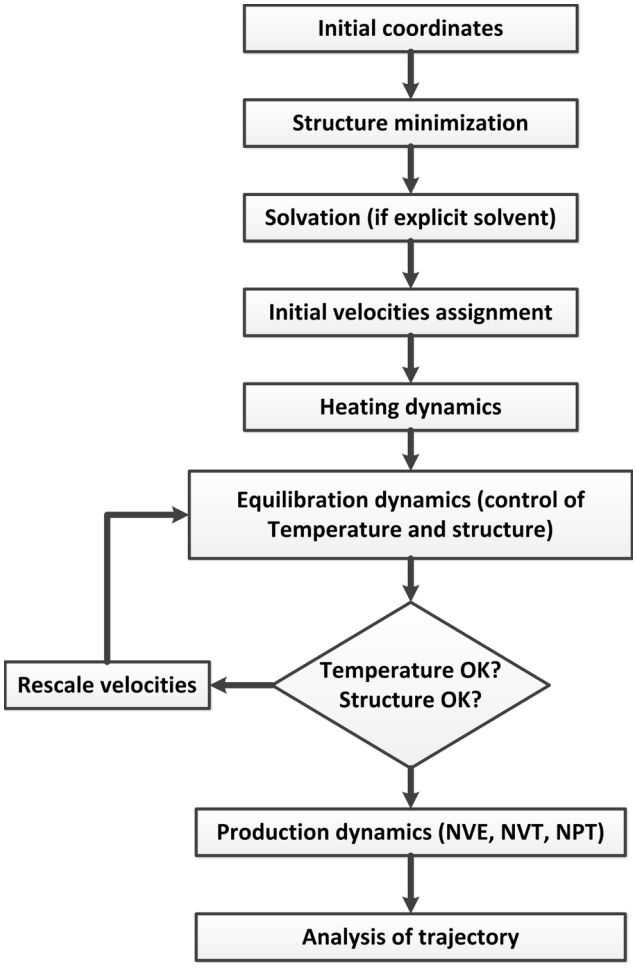

Figure 2.3 The flowchart of MD simulation.

2.3.4.1 Initialization

It can be seen from the previous section that initial positions and velocities of all the atoms need to be specified at the start of the simulations. The initial position can be obtained from experimental or theoretical data such as a homology model, X-ray

crystallography, or NMR structure.

The initial velocities must be assigned to meet two conditions: 1). achieve the desired temperature and 2). the total linear momentum of the system is zero. The velocities of the atoms are related to the kinetic energy as

$$K = \sum_{i=1}^{N} \frac{m_i v_i^2}{2} \qquad \text{(Equation 2.37)}$$

The kinetic energy is related to the temperature as

$$\langle K \rangle = \frac{3}{2} N k_B T \qquad \text{(Equation 2.38)}$$

where k_B is the Boltzmann's constant and T is the thermodynamic temperature. Based on the relationship of velocity and temperature, the initial velocities can be assigned based on a given temperature using Maxwell-Boltzmann distribution

$$\rho(v_i) = \left(\frac{m_i}{2\pi k_B T}\right)^2 \exp\left(-\frac{1}{2}\frac{m_i v_i^2}{k_B T}\right) \qquad \text{(Equation 2.39)}$$

The initial velocities can be randomly selected for each of the 3N components. But it should be noted that randomly selecting the initial velocities in this way results in a nonzero total linear momentum of the system. Therefore, the velocities should be shifted to remove the nonezero momentum.

2.3.4.2 Energy minimization

If MD was started immediately, the added hydrogens and broken hydrogen bond network in water would cause very large forces and structure distortion. Therefore, it is necessary to first perform a short energy minimization to remove the strong forces before starting

the simulation. Energy minimization is needed to allow water molecules to adjust to the structure of interest and correct distorted bond angles. The energy of the system is calculated with the force field. The minimization process can alter the conformation of the system to find lower energy conformations. It is impossible to find the true global minimum given the 3N degrees of freedom and huge number of local minima. In practice, there are many minimization algorithms that can be used such as steepest descent, conjugate gradient, and Newton-Raphson. As an iterative approach, the steepest descent algorithm adjusts the step size using the gradient of the potential energy surface to guide toward the local minima. The steepest descent algorithm converges slowly and can be used for highly restrained systems. The conjugate gradient algorithm uses intelligent choices of search direction and is efficient but costly for large systems. The Newton-Raphson algorithm assumes the potential energy is quadratic in the region of the minima and calculates both slope of energy and rate of change to find the minima. It is also costly for a large molecular system.

2.3.4.3 Solvent and periodic boundary conditions

In real cases, a molecule is generally immersed in solvent (usually water) rather than being isolated. Solvation has an essential influence on the molecular conformation, electrostatic interactions, and binding energies, etc. Therefore, it is critical to model the solvation for MD simulation. There are two ways to model the solvation: 1). explicit treatment when solvent molecules are added to the system; 2). implicit treatment when solvent is modeled as a continuum dielectric. Moreover, in order to keep a realistic solvent simulation, periodic boundary conditions are needed to establish the computational boundaries. As shown in **Figure 2.4**, the cubic system is replicated

infinitely in all directions so that the particles can experience forces as if they were in a bulk solution.[181] An atom moving across one boundary comes back into the system on the other side with identical velocity.

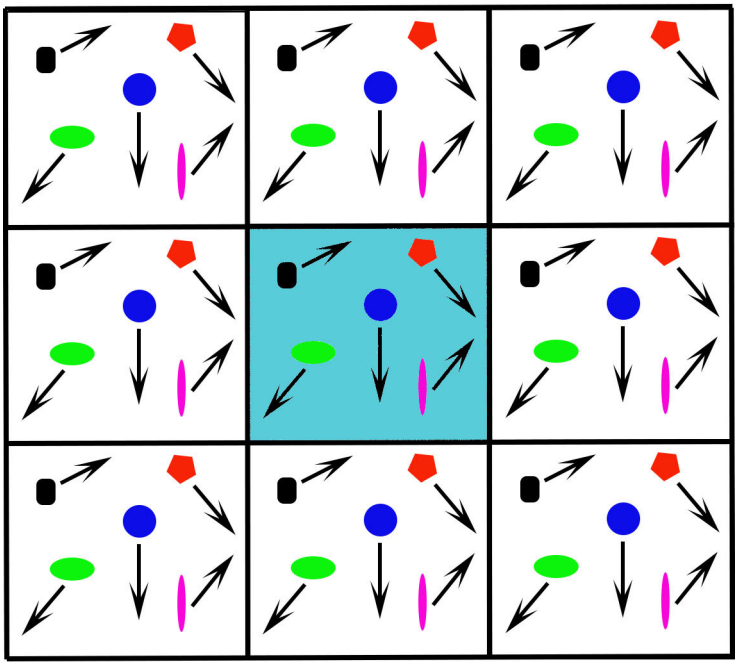

Figure 2.4 Periodic boundary conditions.

2.3.4.4 Heating and equilibration

The initial velocities are generally assigned for a low temperature (for example, 0 K) and the system is heated to gradually increase the temperature. Once the desired temperature is reached, the equilibration procedure is performed and the properties of the system are monitored. The purpose of the equilibration is to run the simulation until the desired pressure, volume, temperature, and energy properties become stable. There are numerical temperature control (thermostats) methods. The simplest way to modify the temperature is to do velocity scaling. Combining Equation 2.37 and 2.38, the following velocity scaling factor can be obtained:

$$\lambda = \sqrt{T_W / T(t)} \qquad \text{(Equation 2.40)}$$

where T_W is the desired temperature and $T(t)$ is the kinetic temperature at time t. The new velocities becomes $v_i^{new} = \lambda v_i^{current}$.

2.3.4.5 Ensembles

An ensemble is a collection of microscopic states that are described by state variables, such as energy E, volume V, temperature T, pressure P, chemical potential μ, and number of particles N. In general, certain state variables need to be remain constant during the simulation to mimic experimental conditions. Depending on which state variables are kept fixed, there are different statistic ensembles that can be obtained as shown in **Table 2.3**. NVT and NPT are the commonly used ensembles in MD simulations. The NVT ensemble can be obtained by temperature-bath coupling during the data collection phase and the volume is maintained constant during the simulation[182]. The NPT ensemble requires control over both the temperature and pressure. The pressure is controlled by adjusting the volume. The NPT ensemble can also be used during the equilibration phase to acquire desired temperature and pressure before changing to other ensembles.

Table 2.3 Example of ensembles

Ensembles	Fixed Variables
Microcanonical (NVE)	N, V, E
Canonical (NVT)	N, V, T
Constant P-T (NPT)	N, P, T
Grand Canonical	μ, P, T

2.3.4.6 Production and analysis

After the equilibrium state is reached, the simulation can be performed for the desired

time length, which is known as the "production" phase. During the production phase, the thermodynamic parameters are calculated and the simulation is conformed to the desired ensemble. Finally, the obtained results (including coordinates, velocities, and other properties) can be used for analysis.

2.4 Analysis

After the MD simulations are performed, the time dependent trajectories, velocities, forces and other properties are obtained and saved. The outputs of MD simulations can then be used for analysis, where information related to conformation, motions, rotations, interactions, and energies of the system can be evaluated. The following methods can be used to analyze the MD simulation results in a time dependent way.

2.4.1 Root mean square deviation and fluctuation

The root mean square deviation (RMSD) has been widely used to measure the difference between structures with respect to a reference structure and is defined by

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i(t) - r_i^{ref}\right)^2} \qquad \text{(Equation 2.41)}$$

where N is the number of atoms and $r_i(t)$ indicates the position of atom i at time t. It should be noted that least square fitting of the two structures is needed to find the corresponding atom pairs between the two structures before the RMSD calculation. The fitting procedure does not have to use the same atoms. The RMSD can be calculated for both the backbone and for the whole protein. In general, the starting structure can be used as reference structure. If protein folding is studied, the folded structure from experiment

can also be used as reference.

The root mean square fluctuation (RMSF) is a measure of the deviation between the position of atom i and some reference position. It is computed by

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}\left(r_i(t_j) - r_i^{ref}\right)^2} \qquad \text{(Equation 2.42)}$$

where T is the time span of the averaging and r_i^{ref} is the reference position of atom i, which is typically defined by the time-averaged position of the same particle (i.e., $r_i^{ref} = \langle r_i \rangle$). Comparing Equation 2.41 and 2.42, we can see that RMSF is averaged over time, generating value for every atom. RMSD, on the other hand, is averaged over the atoms, generating time specific values.

2.4.2 Radius of gyration

The radius of gyration (Rg) is a measure of the compactness of a structure and it is defined by

$$R_g = \sqrt{\frac{\sum_i (\boldsymbol{r}_i - \boldsymbol{r}_c)^2 m_i}{\sum_i m_i}} \qquad \text{(Equation 2.43)}$$

where m_i is the mass of particle i, $\boldsymbol{r}_i$ and $\boldsymbol{r}_c$ are the position of the particle i and the position of the center of mass of the structure. $\boldsymbol{r}_i - \boldsymbol{r}_c$ represents the distance between the particle and the center of mass. Rg can be used to evaluate the shape change of the structure during the simulations. For example, during protein folding, Rg gradually decreases until becoming steady when the protein is folded.

2.4.3 Hydrogen bonds analysis

In biomolecular systems, hydrogen bonds are an important non-covalent structural interaction and are determined by both distance (Donor-Acceptor) and angle (Hydrogen-Donor-Acceptor). The general rule of thumb for determining whether a hydrogen bond exists is based on geometrical criteria: 1). the distance between acceptor and hydrogen is no larger than a distance (e.g., 0.35 nm). 2). the angle between Hydrogen-Donor-Acceptor is no larger than an angle (e.g., 20^{o}). In order to calculate the lifetime of the hydrogen bonds, the time autocorrelation function needs to be calculated first as

$$C(\tau) = \langle h_i(t) h_i(t+\tau) \rangle \qquad \text{(Equation 2.44)}$$

where $h_i(t) = \{0,1\}$ is the existence function of the hydrogen bond i at time t. The relevant hydrogen bond lifetime can then be found via integration of the autocorrelation function as

$$\tau_{HB} = \int_0^\infty C(\tau) d\tau \qquad \text{(Equation 2.45)}$$

2.4.4 Distance analysis

The distances in the structure can also be analyzed to obtain information about contacts in the structure. Both the distances between two particles and the minimum distance between two groups of atoms can be calculated and plotted. The distance between the geometrical centers of two groups can be calculated, so is the minimum distance between two groups of atoms during time. A symmetric matrix of contacts can also be plotted if the minimum distances between all residuals (the smallest distance between any pair of

atoms of the two residues) of the protein are calculated. Analyzing the change of the matrices in time can provide information about the changes of the structure in time.

2.4.5 Principal component analysis (PCA)

Principal component analysis (PCA) is a statistical technique that can transform a set of potentially coordinated observations into a set of linearly uncorrelated variables using orthogonal vectors known as principal components (PCs).[183-186] PCA can be used to reduce the dimensionality of the data to a limited number of PCs. The PC describes the directions in which large variances of coordinates occur. The first principal component has the largest variance, which represents the dominant motion. Therefore, the first few PCs represent correlated modes of motion that occurred during the simulations.

The first step of PCA is an alignment process that removes translation and rotation (i.e., removing six degrees of freedom) of the entire molecule for every frame in the trajectory by least square fit of each frame to a reference. Once the structure has been aligned, the covariance matrix can be calculated. For a given trajectory of N atoms, the $3N \times m$ coordinate matrix can be expressed as $R_i(t) = (x_i(t), y_i(t), z_i(t))$, where $i = 1,2, \dots N$ and $t = 1,2, \dots m$ with m equal to the duration of the trajectory. For simplicity, let $r(t)$ represents any one of $x_i(t), y_i(t),$ or $z_i(t)$. The mass-weighted covariance matrix $\boldsymbol{C}$ can be calculated as

$$C_{ij} = \langle M_i^{\frac{1}{2}} M_j^{\frac{1}{2}} (r_i(t) - \langle r_i(t) \rangle)\big(r_j(t) - \langle r_j(t) \rangle\big) \rangle \qquad \text{(Equation 2.46)}$$

where C_{ij} represent the covariance between ith and jth atoms. M_i represents the mass. $\langle r_i(t) \rangle$ is the time average position along the entire trajectory. It should be noted that

$r_i(t)$ can represent any one of $x_i(t), y_i(t)$, or $z_i(t)$, which means the covariance matrix is a $3N \times 3N$ matrix with $3N - 6$ degrees of freedom.

Then the covariance matrix is diagonalized as

$$\boldsymbol{C} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^{-1} \quad \text{(Equation 2.47)}$$

where V is a square matrix whose columns are eigenvectors and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. After diagonalization, we can obtain $3N$ eigenvectors and eigenvalues, which describe the modes of the collective motion and their respective amplitudes. A eigenvector $\boldsymbol{v_i}$ and corresponding eigenvalue λ_i satisfies the linear equation

$$\boldsymbol{C}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i \quad \text{(Equation 2.48)}$$

The eigenvectors can be sorted in descending eigenvalue index (i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{3N}$) and the first PC can be calculated by projecting the data onto the eigenvector ($\boldsymbol{v_1}$) corresponding to the largest eigenvalue as

$$\boldsymbol{P}_1(t) = \boldsymbol{v}_1 \cdot \boldsymbol{r}(t) \quad \text{(Equation 2.49)}$$

where $\boldsymbol{r}(t) = (r_1(t), r_2(t), \ldots, r_{3N}(t))^T$. Similarly, the second PC, $\boldsymbol{P}_2(t)$, corresponding to the second largest eigenvalue can be obtained and so forth. The trajectory can be projected onto each eigenvector for all time steps and followed by investigating the time evolution of the projected trajectory. The resulting distribution of each projection has variance of λ_i, which is the physical meaning of the eigenvalues.

2.4.6 Free energy landscape

In analysis of MD simulations, the free energy landscape (FEL) can be used to map the possible conformations of a molecule during the simulation along with the corresponding Gibbs free energy, which can be achieved by an appropriate conformational sampling procedure. FEL shows the hills and valleys of free energy that a molecule adopted during the simulation, which can be useful in the selection of representatives to use in further analysis. In general, FEL can be represented with two variables that reflect specific properties of the system and free energy as the third variable, which can be estimated from the probability distributions of the system. For example, the first two variables can be the first and second principal components of the protein after PCA. Then the FEL shows valleys of low energy that represent metastable conformations of the system. The energy minimum structures during the simulations can be identified. The free energy landscape is calculated as:

$$G_{(PC1,PC2)} = -k_B T ln P_{(PC1,PC2)} \qquad \text{(Equation 2.50)}$$

where k_B is the Boltzmann constant, T is the temperature, and $P_{(PC1,PC2)}$ represents the normalized joint probability distribution.

2.4.7 Residue interaction network and WebPSN

Residue interaction network (RIN) is a technique that can be used to analyze the protein structures.[187-189] RIN can represent protein structures with network nodes and arcs, where nodes represent residues and arcs that links the nodes indicate interactions between residues. Using the network theory to represent protein structures can significantly

simplify the structure analysis and enable the users to focus on the specific interested residues. In this study, RING-2 is used for RIN analysis due to its various capabilities, such as to calculate hydrogen bond, van der Waals, secondary structure, distance-based generic contacts, intra- and inter-chain interactions, and contacts with ligands.[189] First, RING-2 uses the measured atom distances to create a list of residue-residue (or residue-ligand or ligand-ligand) pairs that are capable of interaction, followed by methods to identify interaction types and create attributes for the nodes and edges. RING-2 uses geometrical criteria to estimate atomic interactions instead of using a complicated force field analysis, which significantly reduces the processing time. It has been found that RING-2 has satisfying performance in analyzing mutation effects, protein folding, protein domain-domain communication, and catalytic activity.[189]

In addition to RIN, a free web server WebPSN is also been used in this study to analyze the structurally important residues, protein stability, and communication.[190] WebPSN is a mixture of protein structure network (PSN) and elastic network model-normal model analysis (ENM-NMA). PSN analysis has been used to analyze the protein structures, functions, dynamics, communications, and folding processes.[191,192] PSN also creates a network using amino acid residues as nodes and non-covalent interactions between residues as the corresponding edges. The strength of the interaction between two residues i and j can also be calculated in percentage as

$$I_{ij} = \frac{n_{ij}}{\sqrt{N_i N_j}} \times 100 \qquad \text{(Equation 2.51)}$$

where n_{ij} represents the number of atom-atom pairs between the side chains of residues

within a distance cutoff. N_i and N_j are the normalization factors. NMA is a method to study the structural and dynamic properties of protein systems and predicts functional motions. ENM-NMA is a coarse grained NMA method that can describe the vibrational dynamics of complex systems.[193]

In WebPSN, the PSN analysis is first implemented on a single high-resolution structure and a protein structure graph is generated. The shortest communication pathways on ensembles of structures are generated by defining all possible communication paths between selected node pairs, followed by a filtering procedure based on cross-correlation of atomic motions.

2.4.8 Binding free energy calculation

The Molecular Mechanics Poisson-Boltzmann Surface Area (MM-PBSA) method is a reliable and efficient binding free energy calculation method that has been widely used in MD simulations to model protein-ligand binding interactions. MM-PBSA combines continuum electrostatic solvation with a ligand-receptor van der Waals energy, intramolecular, stereochemical energy terms, a nonpolar solvation term proportional to buried surface area, and vibrational entropies.[194-200] In MM-PBSA, the binding free energy of the protein-ligand complex in an aqueous solvent can be decomposed into the relative free energy of the solvated receptor-ligand complex and the separated, solvated ligand and receptor as

$$\Delta G_{\text{bind,solv}} = \Delta G_{\text{complex,solv}} - \Delta G_{\text{receptor,solv}} - \Delta G_{\text{ligand,solv}} \quad \text{(Equation 2.52)}$$

Each of the free energy change term in Equation 2.52 can be decomposed as:

$$\Delta G_{\text{solv}} = \Delta E_{\text{MM}} + \Delta G_{\text{bind,solv}} - T\Delta S \qquad \text{(Equation 2.53)}$$

where ΔE_{MM}, $\Delta G_{\text{bind,solv}}$, and $-T\Delta S$ represent the vacuum-phase molecular mechanical energy change, solvation free energy change, and the entropy change upon binding, respectively. The molecular mechanical energy and solvation free energy change can be further decomposed as

$$\Delta E_{\text{MM}} = \Delta E_{\text{covalent}} + \Delta E_{\text{electrostatic}} + \Delta E_{\text{vdW}} \qquad \text{(Equation 2.54)}$$

$$\Delta G_{\text{bind,solv}} = \Delta G_{\text{polar}} + \Delta G_{\text{non-polar}} \qquad \text{(Equation 2.55)}$$

where $\Delta E_{\text{covalent}}$, $\Delta E_{\text{electrostatic}}$, and ΔE_{vdW} represent the covalent energy change, the electrostatic energy change, and the van der Waals energy change, respectively. ΔG_{polar} and $\Delta G_{\text{non-polar}}$ represent polar and non-polar contributions, respectively. The covalent energy change contains of changes in the bond terms (ΔE_{bond}), the angle terms (ΔE_{angle}), and the torsion terms ($\Delta E_{\text{torsion}}$) as

$$\Delta E_{\text{covalent}} = \Delta E_{\text{bond}} + \Delta E_{\text{angle}} + \Delta E_{\text{torsion}}. \qquad \text{(Equation 2.56)}$$

All of the above energy changes can be calculated by averaging over a large set of sampled conformations.[200] The entropy term is the most difficult to compute and has the least accuracy; it can be approximated with a normal mode analysis or a quasi-harmonic analysis.[201] The polar solvation term can be calculated with a numerical solution of the Poisson-Boltzmann equation. The non-polar solvation free energy can simply be estimated to be proportional to the solvent accessible surface area (SASA) of the solute:

$$\Delta G_{\text{non-polar}} = \gamma * \text{SASA} + b \qquad \text{(Equation 2.57)}$$

where γ is the surface tension and b is the correction term. Both terms are usually set to be constant for all solute molecules.[200] For example, γ=0.00524 kcal/mol/Å^2 and $b = 0.92$ kcal/mol.[201]

Chapter 3 Investigating the substrate binding mode and binding affinity of quorum-quenching acylase PvdQ with multiple docking strategies

3.1 Introduction

P. aeruginosa is one of the most common Gram-negative opportunistic pathogens associated with hospital stays. People with cystic fibrosis, severe burns, AIDS, or other immuno-compromised patients are susceptible populations for this bacterial infection.[202,203] *P. aeruginosa* is difficult to eradicate due to its capability to form a biofilm, which can significantly increase the resistance to the common antibiotics.[204] Their co-colonization with other bacteria can exacerbate the disease and lead to a poorer prognosis. *B. cenocepacia* is found co-colonized with *P. aeruginosa* in patients with lung infection. *Bcc* is a group of Gram-negative opportunistic human pathogens, which can cause pneumonia in people with compromised immune systems along with underlying lung disease. More and more studies have been focusing on these bacteria due to the seriousness of infection as well as the difficulty to eradicate them with common antibiotic if they are co-infected with other bacteria.[204,205] Therefore, it is critical to find new ways to inhibit the bacterial virulence and biofilm formation without adding more selective pressure.

It is well-known that bacteria are communicative microorganisms and thereby are able to organize and coordinate their behavior through QS. Both *P. aeruginosa* and *B.*

cenocepacia employ QS to control their actions. *P. aeruginosa* has a well-known QS system. 3-oxo-C12-HSL and C4-HSL are the two major endogenously generated communication molecules, mediating the QS.[206] C8-HSL is the signal molecule for *B. cenocepacia*. The degradation of HSL signaling molecules is an attractive way to inhibit QS and target bacterial infections.[205] One of the structurally well-characterized quorum quenching enzymes is AHL acylase PvdQ, which can break down the amide bond between the AHL head group and the acyl tail group. According to previous studies, PvdQ has different preference for different substrates.[207,208] It prefers to target the AHLs with long acyl chains, such as 12 or 14 carbons in tail group, rather than AHLs with short tail groups or 3-oxo substituents, such as C8-HSL, C4-HSL, and 3-oxo-C12-HSL. The determinants of substrate recognition and discrimination of PvdQ with different bacterial ligands are still unknown. Therefore, it is necessary to use other methods to study the determinants of the substrate binding specificity of PvdQ using native AHL ligands. The current work addresses this problem by using multiple docking approaches to investigate the binding mode and binding affinity of PvdQ protein.

Molecular docking calculation is an essential way to study the biological activities and correlations of protein and ligand structures. Reliable and accurate prediction of the binding modes and binding affinity of protein bound with a ligand is critical to study their function in many fields.[209] Various programs are currently available for prediction of docked poses. As addressed in Wang's paper, no single docking program has superiority over the others. Generally, combining different docking methods appears to be a practical method for achieving better predictions.[165] Therefore, three of the most frequently used protein-ligand docking programs (i.e., Schrödinger IFD, GOLD, and AutoDock Vina)

were employed in this study to obtain more reliable binding modes and binding affinities of different ligands (C12-HSL, 3-oxo-C12-HSL, C8-HSL, and C4-HSL) bound to the PvdQ protein. The combination of multiple molecular docking approaches in this study can not only be applied to accurately predict the binding poses and binding affinities of PvdQ with different substrates, but also provide starting structures for further MD simulation.

3.2 Computational methods

3.2.1 pKa prediction

The University of Houston Brownian Dynamics (UHBD) program[152] was used to perform pKa predictions in order to predict the ionization states of key residues in the PvdQ protein. All of the non-protein components were removed before the calculation. pKa's were predicted using the CHARMM force field[210] and the resultant ionized states of the ionizable residues in the protein were selected at pH 7.0. The dielectric constant of the solvent was set to 80, while the dielectric constant of the protein was set to 20. The maximal number of iterations is 300 and the temperature is 293 K. The ionic strength and radius of ions were 150 mM and 2.0 Å, respectively. The grid spacing and dimensions were set to four different levels of focusing [4: 2.5 45 45 45; 1.2 15 15 15; 0.75 15 15 15; 0.25 20 20 20].

3.2.2 Schrödinger IFD

The Schrödinger IFD program was used to investigate the interactions between PvdQ protein and different ligands.[166] IFD employed an initial Glide docking with subsequent

Prime induced fit docking to take the flexibility of both protein and ligand into account (**Table 3.1**). The protein was prepared with the protein preparation wizard in the Maestro visualizer. All waters were removed, all hydrogens were added to the protein, and bond orders were assigned. All hydrogen-bonding networks were optimized and the ionization states were assigned at pH 7.0. The structure of the protein was optimized and minimized using the OPLS_2005 force field before docking. The ligand was prepared with the LigPrep platform in the Maestro visualizer and was energy minimized using the OPLS_2005 force field. The ligand was treated as fully flexible. The docking grid was generated with the Receptor Grid Generation tool. The default van der Waals radius scaling factors were used. The center and size of the grid box were defined according to the position of the hydrolysis product, 3-oxo-lauric acid, in complex with PvdQ in the crystal structure (PDB ID: 2WYC). The size of the grid box was 30 Å. The ligand was docked into the receptor via initial Glide docking to generate a maximum of up to 50 poses with a default van der Waals scaling factor; then the docked poses were subjected to the Prime induced fit docking. The residues within 5 Å of each ligand pose were energy-minimized and refined to make the substrate-binding pocket flexible. The extra precision docking score was used to rank the docking poses. No constraints were applied and all the other settings remained at their defaults.[209]

Table 3.1 The settings of the Schrödinger IFD

Schrödinger IFD	
Receptor	The residues within 5 Å of each ligand pose are flexible
Ligand	Flexible
Force field	OPLS_2005 force field
Score functions	Extra-Precision
Hydrogen	Add hydrogen to both receptor and ligand
pH	7.0
Gridbox center	Defined according to the position of the hydrolysis product of PvdQ complex (PDB ID: 2WYC)
Gridbox range	30 Å × 30 Å × 30 Å

3.2.3 GOLD docking

The GOLD Suite 5.4.1 (Genetic Optimization for Ligand Docking; Cambridge Crystallographic Data Centre)[158,211] docking method was also used to investigate the binding mode and binding affinity of PvdQ with different ligands. GOLD employed pharmacophore point matching and a genetic algorithm (GA) to generate the binding poses and calculate binding affinities. The protein and ligand were prepared using the Hermes visualizer in the GOLD suite (**Table 3.2**). Briefly, all hydrogens were added to the protein. The center of the binding site (X, Y, Z: 35.242, 52.079, 49.826) was defined by the position of the substrate in the 2WYC structure. All atoms within a radius of 15 Å of the binding ligand were defined as the binding site. The side chain of the key active residues Serβ1 (catalytic center), Pheβ24 (acting as a gate), Valβ70 and Asnβ269 (oxyanion hole residues), Trpβ186 and Leuα146 (residues that move upon ligand binding) were treated as flexible during the whole docking calculation to improve the fitness of the ligand to the protein. The ligand was also treated as being fully flexible. 500

GA runs were selected for the ligand docking calculation. The ChemScore scoring function was used for this docking study. No early termination was allowed. The default GA search options were chosen to obtain 100% search efficiency. Each docking calculation was repeated three times with different seed values.

Table 3.2 The settings of the GOLD Docking

GOLD Docking	
Receptor	The protein is rigid except that residues Serβ1, Pheβ24, Valβ70, Trpβ186, Asnβ269, and Leuα146 are flexible.
Ligand	Flexible
Score functions	ChemScore scoring function
GA runs	500
Hydrogen	Add hydrogen to both receptor and ligand
Early termination	No
Seed	Random seed value, repeat 3 times
Gridbox center	Defined by the position of the substrate in the 2WYC structure.
Radius	15 Å

3.2.4 Autodock Vina docking

Finally, AutoDock Vina, one of the best-known and widely used open-source protein-ligand docking programs, was employed in our study. It is designed to predict the binding interaction of small molecules and a receptor. Vina improves the average accuracy over AutoDock by using a different scoring function and optimization algorithm.[156] The protein PvdQ and four different ligands structures were prepared with Autodock Tools (**Table 3.3**).[155] All non-protein components were deleted. All hydrogens were added and Gasteiger charges[212] were assigned. The side chains of residues Serβ1, Pheβ24, Valβ70, Trpβ186, Asnβ269, and Leuα146 along with the entire ligand were flexible during the

docking procedure. PvdQ was enclosed in a grid box of 30 Å in each direction to include the substrate binding site. The center of the box was the same as the other two docking methods. In order to find the best conformations, the maximum number of binding modes to generate and the level of exhaustiveness were set to 20 and 16, respectively. The docking score was utilized to rank the binding affinity of each docking pose. Each docking calculation was repeated 10 times with different random seeds.

Table 3.3 The settings of the Vina Docking

Vina Docking	
Receptor	The protein is rigid except that residues Serβ1, Pheβ24, Valβ70, Trpβ186, Asnβ269, and Leuα146 are flexible.
Ligand	Flexible
Num_modes	20
Exhaustiveness	16
Seed	Random, repeat 10 times
Charges	Gasteiger charges were assigned
Gridbox center	Defined by the position of the substrate in the 2WYC structure.
Grid spacing	1 Å
Gridbox range	30 Å × 30 Å × 30 Å

3.3 Results and discussion

3.3.1 pKa results

In order to predict the ionization states of important residues in the protein, pKa predictions were carried out to better understand the structure-function relationship and

aspects of the catalytic mechanism of the PvdQ protein. The titration curve of the N-terminus of Serβ1 demonstrates that it exhibits a significant shift toward the acidic range (**Figure 3.1**), which means the N-terminus of Serβ1 is kept in its neutral form. This confirms the proposal from Clevenger's study that the catalytic center Serβ1 can directly deprotonate its own hydroxyl group through its own N-terminal amine.[207] The surrounding residues (i.e., Hisβ23 and electropositive Argβ297) having the highest electrostatic interaction free energy with Serβ1 are responsible for the pKa depression of the N-terminus of Serβ1, which is also consistent with the results from Clevenger's study.[207]

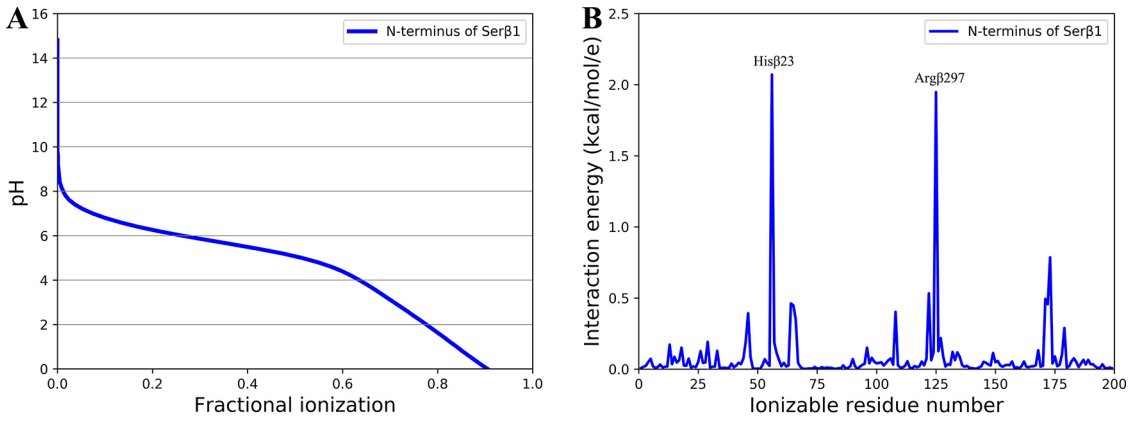

Figure 3.1 (A) The titration curve of the N-terminus of Serβ1 in PvdQ. (B) The electrostatic interaction free energy between the N-terminus of Serβ1 and the other ionizable residues.

3.3.2 Substrate binding modes and binding affinity analysis from Schrödinger IFD

In order to study the binding interactions between PvdQ and different substrates, multiple docking methods were used to obtain reliable docking results. In fact, the side-chains or backbones of many proteins move when bound with other substrates. These movements

can change the binding sites to better accommodate the ligands. Therefore, Schrödinger IFD was adopted to involve the flexibilities of both ligand and receptor. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each ligand, the docking score, and the orientation of the binding poses were used to select biologically reasonable binding poses. As represented in **Table 3.4**, the binding affinities of the most favorable poses of C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL are -9.9 kcal/mol, -8.7 kcal/mol, -7.7 kcal/mol, and -5.9 kcal/mol, respectively. These binding affinities are positively correlated with the *N*-acyl chain length after considering the fact that there are ranges of standard error in each docking method. The distances between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL are 3.0 Å, 5.8 Å, 6.8 Å, and 7.4 Å, respectively. The distances are negatively correlated with the N-acyl chain length, which is consistent with the binding affinities of each protein-ligand complex.

Table 3.4 Distance between the catalytic center Serβ1 Oγ and the carbonyl carbon atom of the amide bond of each ligand, and docking scores obtained from the Schrödinger IFD

Schrödinger IFD results		
Complex Names	Distance (Å)	Docking score (kcal/mol)
C12-HSL	3.0	-9.9
3-oxo-C12-HSL	5.8	-8.7
C8-HSL	6.8	-7.7
C4-HSL	7.4	-5.9

As shown in **Figure 3.2**, the sliced hydrophobicity surfaces of PvdQ binding with different substrates are generated to study the protein-ligand interactions. Green represents the most hydrophilic, while purple represents the most hydrophobic (hydrophobic property increases from green to white to purple). The tail group of C12-HSL substrate is buried inside the hydrophobic binding pocket, while the head group is located in the deep solvent-accessible cleft between the two knobs of PvdQ. The orientation of Oγ atom of the catalytic center Serβ1 is close to the carbonyl carbon atom of the amide bond of C12-HSL. This position of C12-HSL is similar to that of the proposed siderophore precursor[207] , which is favorable for the catalysis to occur. The tail group of 3-oxo-C12-HSL is also buried in the hydrophobic binding pocket. However, the head group bends toward the hydrophobic binding pocket. The hydrophilic head group and the extra 3-oxo substituent are surrounded by unfavorable hydrophobic residues. The position of the Oγ atom of the catalytic center Serβ1 is far away from the carbonyl carbon atom of the amide bond of 3-oxo-C12-HSL. Therefore, the catalytic efficiency of 3-oxo-C12-HSL substrate is less than C12-HSL substrate. It should be noted that C8-HSL substrate is totally buried inside the PvdQ protein covered by the whole hydrophobicity surface. The location of Oγ atom of the catalytic center Serβ1 is even furtherer from the carbonyl carbon atom of the amide bond of C8-HSL. Therefore, the catalytic efficiency of the C8-HSL substrate is less than the C12-HSL and 3-oxo-C12-HSL substrates. For C4-HSL, the whole substrate is entirely buried inside the hydrophobic binding pocket. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of C4-HSL is the furthest as compared to other substrates which is consistent with C4-HSL's low catalytic efficiency.

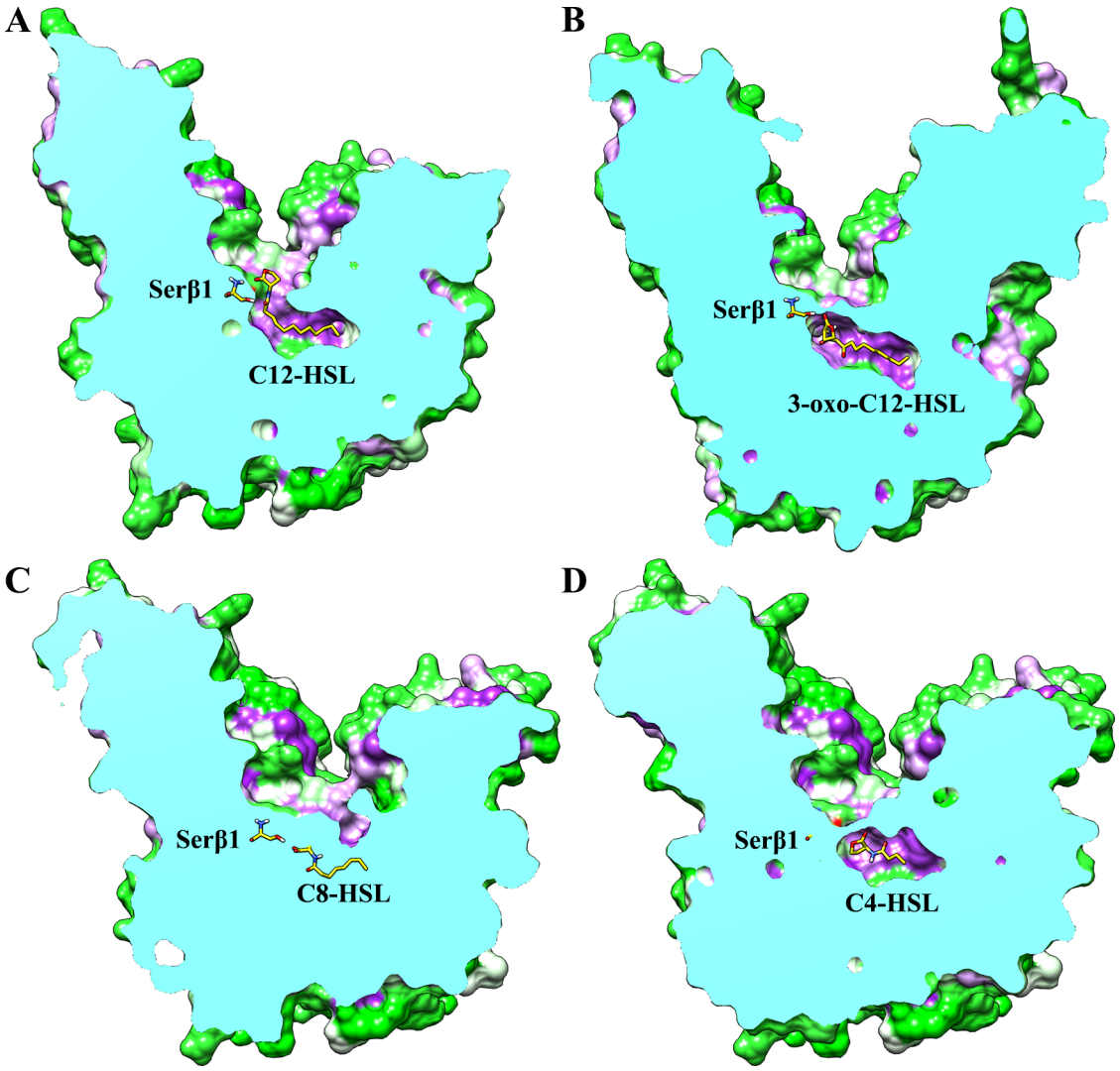

Figure 3.2 Sliced hydrophobicity surface of PvdQ binding with different substrates from the Schrödinger IFD. Green represents the most hydrophilic, while purple represents the most hydrophobic (color changes from green to white to purple). Serβ1 is the catalytic center. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex.

In order to study the detailed interactions of PvdQ with different ligands, the specific docking poses of PvdQ with four different substrates were generated. **Figure 3.3** shows the substrate binding sites of the IFD positions for each PvdQ complex. As shown in **Figure 3.3**, the positions of all ligands after docking are buried inside the binding pockets

of PvdQ. The conformations of the binding pockets have changed slightly to accommodate different ligands. The Oγ atoms of the nucleophile Serβ1 of all complexes are toward the substrate in the binding sites. Residues Pheβ24 of all complexes are in their open states to enable the substrates buried inside the binding pockets. In C12-HSL complex, the Pheβ24 shows slightly bigger twist to fully open the binding pocket as compared to the other ligands. There are no substantial conformational changes of the side-chains of residues Valβ70 and Asnβ269. The movements of the residue Trpβ186 and Leuα146 affect the sizes of the binding pockets. In C4-HSL complex, different side-chain orientations of residues Trpβ186 and Leuα146 reduce the volume of the substrate binding pockets. In this study, we assume that the catalysis is more likely to occur when the distance becomes shorter. For C12-HSL, the distances between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of C12-HSL is 3.0 Å. The most reasonable poses for C12-HSL obtained from Schrödinger IFD methods demonstrates that it has the shortest distance and the most favorable docking score. Therefore, there is greater opportunity for catalysis to occur. The distances for 3-oxo-C12-HSL, C8-HSL and C4-HSL are 5.8 Å, 6.8 Å, and 7.4 Å, respectively. Accordingly, the poses for 3-oxo-C12-HSL and C8-HSL are less energetically favorable in comparison to the C12-HSL complex, considering their distant positions and low docking scores. The location of C4-HSL in the pose is far beyond the hydrogen bonding capability and has the least favorable docking score, which is consistent with its low catalytic efficiency.

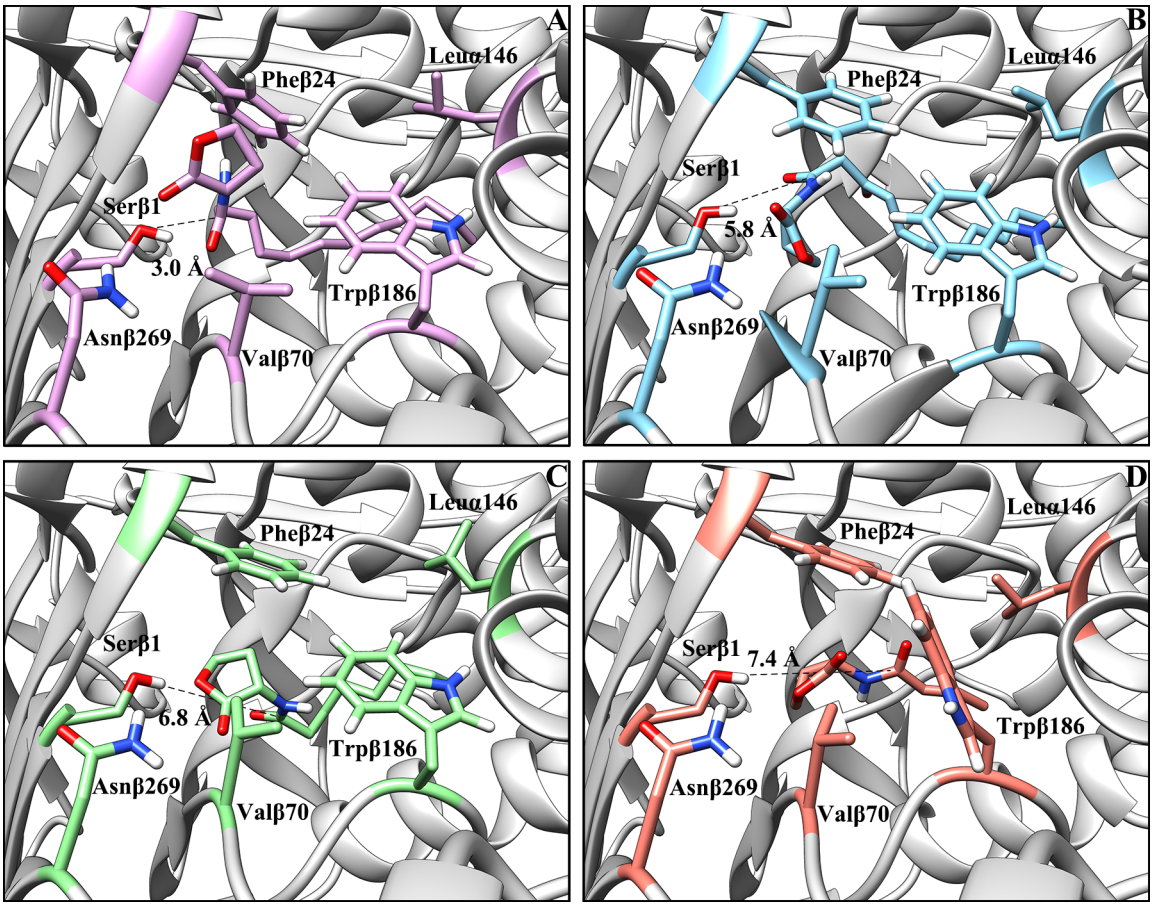

Figure 3.3 The docking poses of PvdQ with four different substrates from the Schrödinger IFD. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex. Residues treated flexible during docking are displayed in stick. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each ligand was measured.

In order to further investigate the hydrogen bonds and hydrophobic interactions of PvdQ with each substrate, the schematic diagrams of protein-ligand interactions were generated by LIGPLOT v.4.5.3.[213] The hydrogen bonds are represented by dashed lines, while the hydrophobic interactions are indicated by arcs with spokes. As shown in **Figure 3.4**, C12-HSL ligand forms three hydrogen bonds with Serβ1 (atom N and atom Oγ) and Hisβ23 from PvdQ protein. Serβ1 is the nucleophilic catalytic residue, which can deprotonate its own hydroxyl group to activate the catalytic role. The close interaction

between residue Serβ1 and C12-HSL indicates a higher chance for catalysis to occur. Hisβ23 forms polar interaction with ligand to stabilize the ligand in an appropriate position. However, there are fewer hydrogen bonds formed between PvdQ and the other ligands as compared to the C12-HSL ligand. These results from Schrödinger IFD program are consistent with the experimental conclusion that the PvdQ protein prefers HSLs with long acyl chains and without the 3-oxo substituent.[205,206]

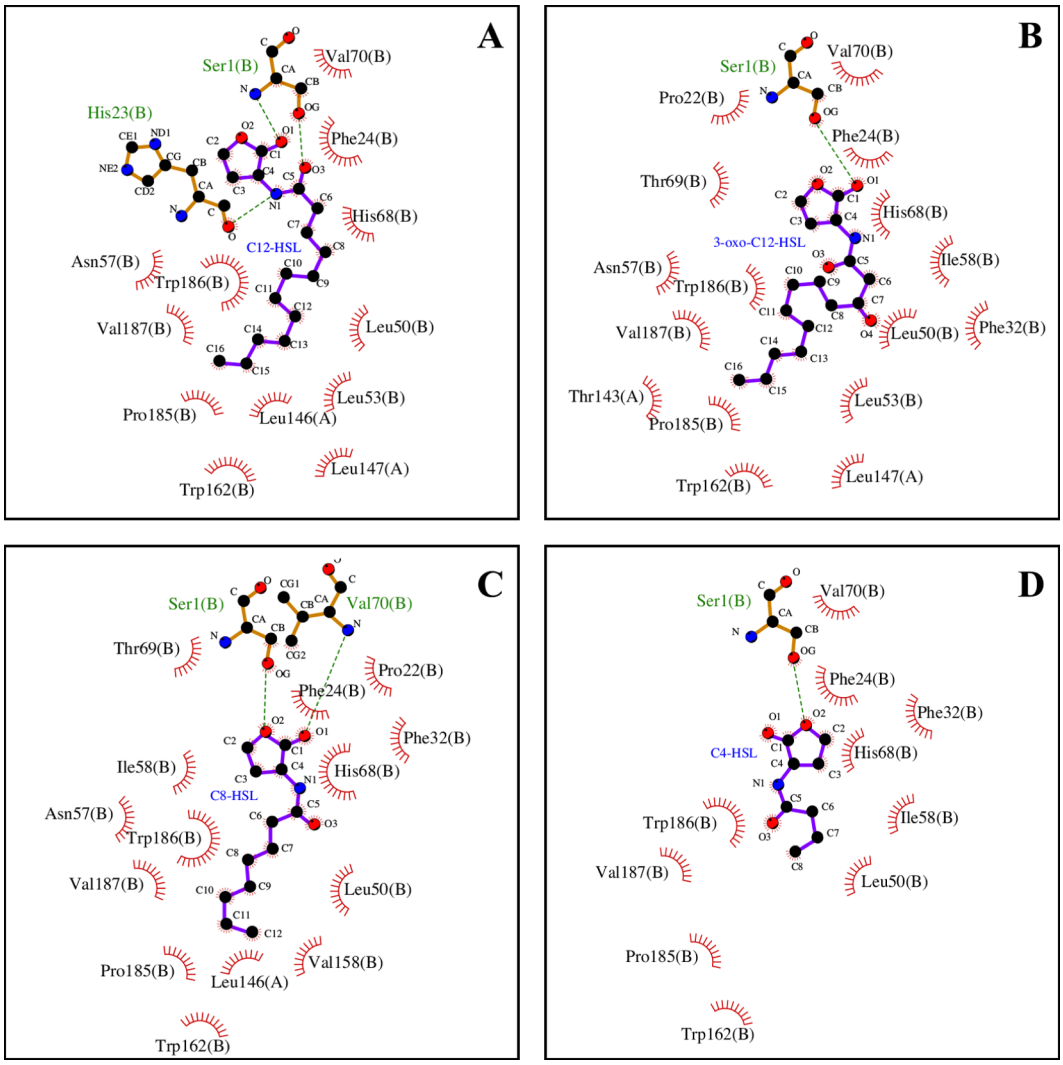

Figure 3.4 The ligand interactions of PvdQ with four different substrates from Schrödinger IFD. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex. The hydrogen bonds are represented by dashed lines, while the hydrophobic interactions are indicated by arcs with spokes.

3.3.3 Substrate binding modes and binding affinity analysis from GOLD docking

GOLD docking program is one of the most robust programs on pose predictions and it possesses high consistent rates of 82.5% (consistent rate is defined as SR_{tsp}/SR_{bp}, where SR_{tsp} is the success rate for the top scored poses and SR_{bp} is the success rates for the best poses).[165] In order to improve the fitness of the ligand to the PvdQ protein, the side chains of residues Serβ1, Pheβ24, Valβ70, Trpβ186, Asnβ269, and Leuα146 along with the entire ligand treated flexible during the docking calculation. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each ligand, the docking score, and the orientation of the binding poses were used to select biologically reasonable binding poses. As illustrated in **Table 3.5**, the fitness scores of C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL are -7.5 kcal/mol, -6.5 kcal/mol, -5.7 kcal/mol, and -5.3 kcal/mol, respectively. The distances between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL are 3.0 Å, 4.9 Å, 3.1 Å, and 7.5 Å, respectively. The shorter the distance, the higher chance for the catalysis to occur. Therefore, we can see that the binding affinities are positively correlated with the *N*-acyl chain length.

Table 3.5 Distance between the catalytic center Serβ1 Oγ and the carbonyl carbon atom of the amide bond of each ligand, and docking scores obtained from the GOLD docking

GOLD docking results		
Complex Names	Distance (Å)	Docking score (kcal/mol)
C12-HSL	3.0	-7.5
3-oxo-C12-HSL	4.9	-6.5
C8-HSL	3.1	-5.7
C4-HSL	7.5	-5.3

As represented in **Figure 3.5,** the sliced hydrophobicity surfaces of different PvdQ-ligand complexes are achieved. The tail groups of C12-HSL, 3-oxo-C12-HSL, and C8-HSL are all buried inside the hydrophobic binding pocket with different depths. The head groups of C12-HSL, 3-oxo-C12-HSL, and C8-HSL are all located in the deep solvent-accessible cleft of PvdQ. The orientations of the substrates decrease the distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each substrate. The tail groups with different acyl chain lengths and extra 3-oxo substituent have different hydrophobic interactions with the residues lining the hydrophobic binding pocket. The whole C4-HSL ligand is buried inside the PvdQ covered with the hydrophobicity surfaces. The Oγ atom of the catalytic center Serβ1 is far from the carbonyl carbon atom of the amide bond of C4-HSL. The totally different interaction pattern of C4-HSL indicates the low catalytic efficiency.

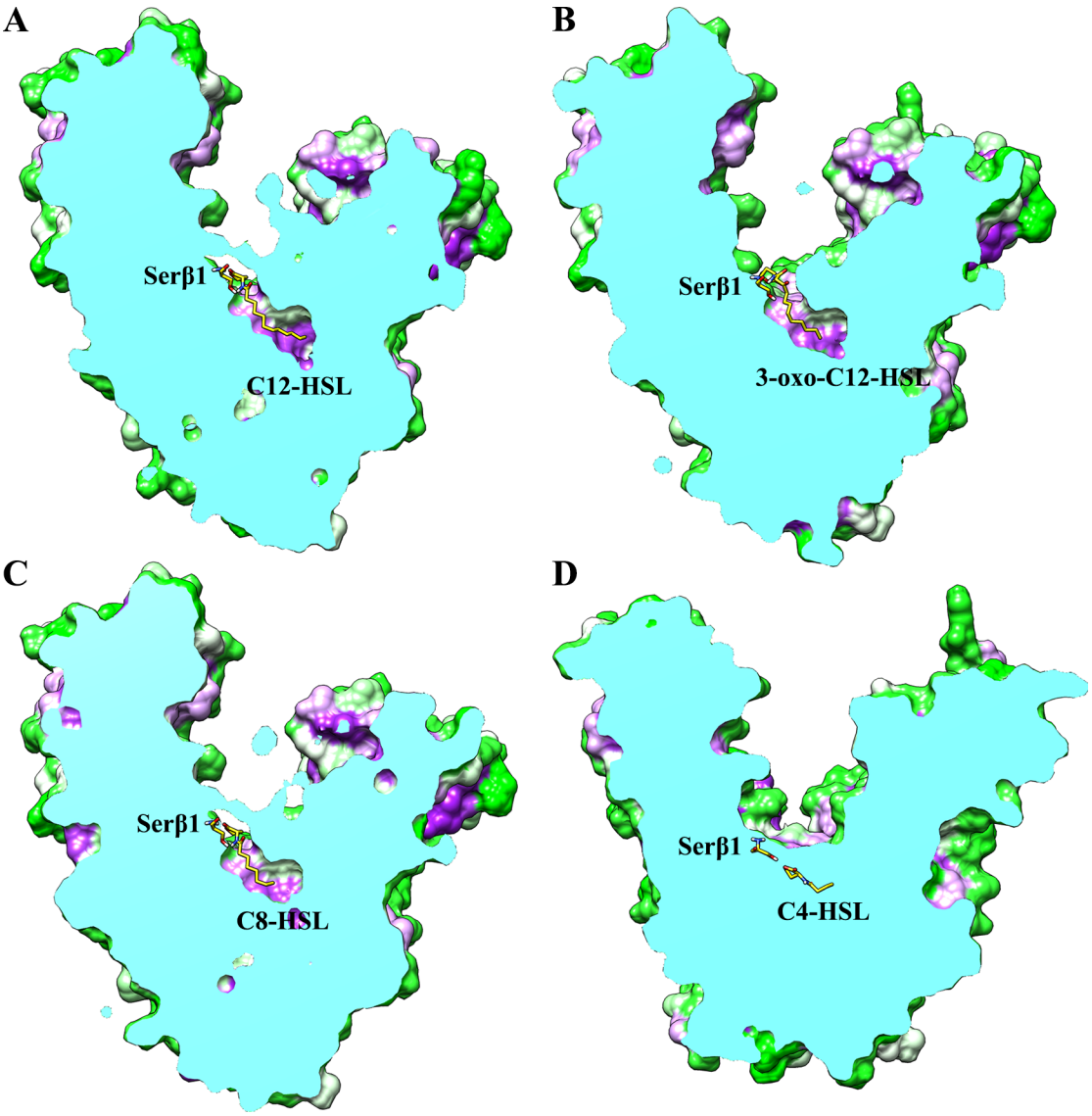

Figure 3.5 Sliced hydrophobicity surface of PvdQ binding with different substrates from the GOLD docking. Green represents the most hydrophilic, while purple represents the most hydrophobic (color changes from green to white to purple). Serβ1 is the catalytic center. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex.

Figure 3.6 shows the binding site of each substrate-bound PvdQ complex from GOLD docking method. As shown in **Figure 3.6,** the pose of each ligand after docking is buried inside PvdQ protein without causing substantial conformational changes in the substrate

binding site. The orientation of the Oγ atom of the catalytic center Serβ1 of each complex is toward the substrate in the binding site. Substantial conformational changes of the side-chain of residues Valβ70 and Asnβ269 are not observed in C12-HSL, 3-oxo-C12-HSL, and C8-HSL complexes. In C4-HSL complex, the movements of the side-chains of residues Pheβ24, Valβ70, and Trpβ186 have changed the volumes of the substrate-binding pockets as compared to other complexes. In each complex, residue Leuα146 has a slight conformational change to properly accommodate different ligands. The distances between the Oγ atoms of the catalytic center Serβ1 and the carbonyl carbon atoms of the amide bond of C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL are 3.0 Å, 4.9 Å, 3.1 Å, and 7.5 Å, respectively. The pose for C12-HSL obtained from the GOLD docking method demonstrates that it has the shortest distance and the most favorable docking score. The pose for 3-oxo-C12-HSL is not quite as energetically favorable as compared to the C12-HSL complex. The position of C8-HSL ligand is similar to that of C12-HSL. However, the shorter acyl tail group lowers the binding affinity. The distance for the C4-HSL pose is far beyond the hydrogen bonding capability and has the least favorable docking score.

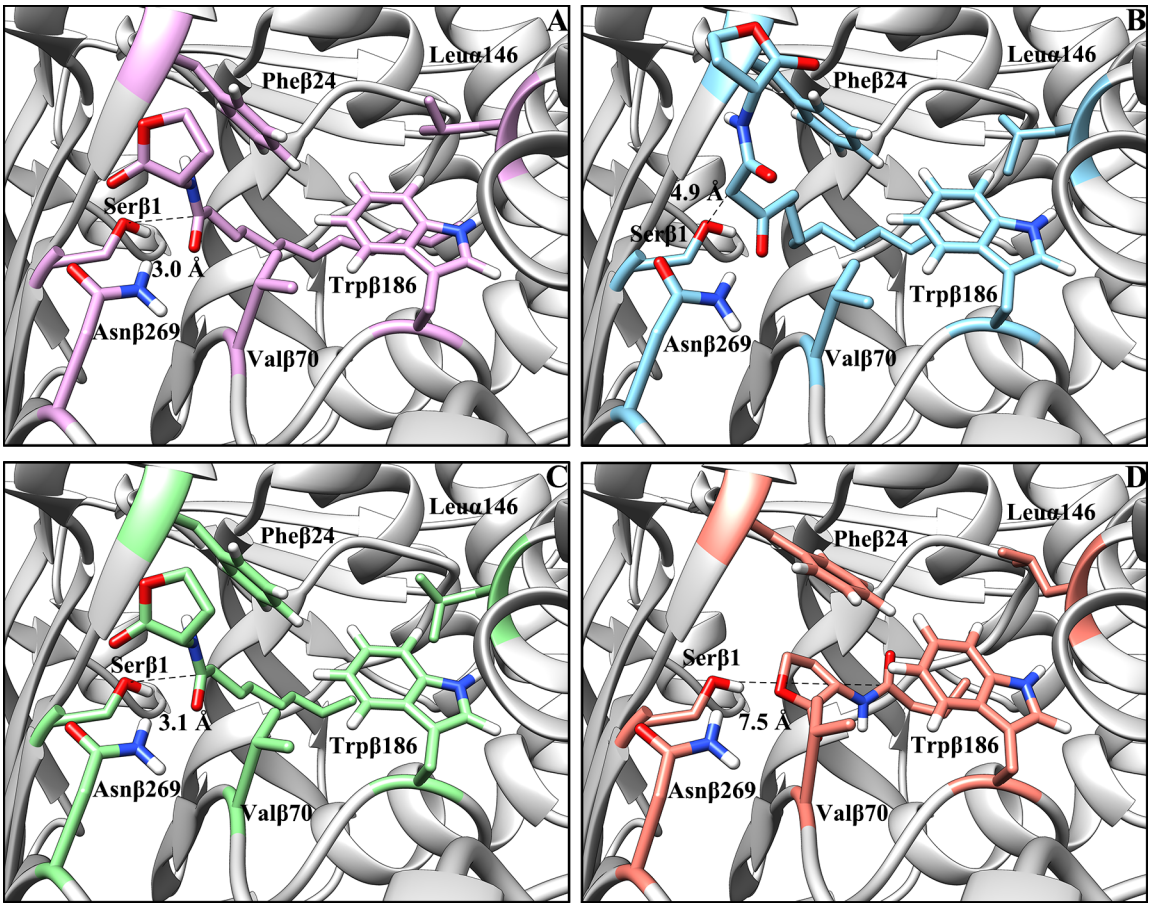

Figure 3.6 The docking poses of PvdQ with four different substrates from the GOLD docking. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex. Residues treated flexible during docking are displayed in stick. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each ligand was measured.

As shown in **Figure 3.7**, there are three hydrogen bonds formed between the C12-HSL ligand and residues Serβ1, Hisβ23, and Valβ70 of the PvdQ protein. C8-HSL forms the same hydrogen bonds with the C12-HSL ligand due to the same position of the head group. 3-oxo-C12-HSL forms three hydrogen bonds with Hisβ23, Valβ70, and Asnβ269, while only one hydrogen bond is formed with the C4-HSL ligand. Basically, these results

from the GOLD docking program are also consistent with the previous experimental results.[205,207]

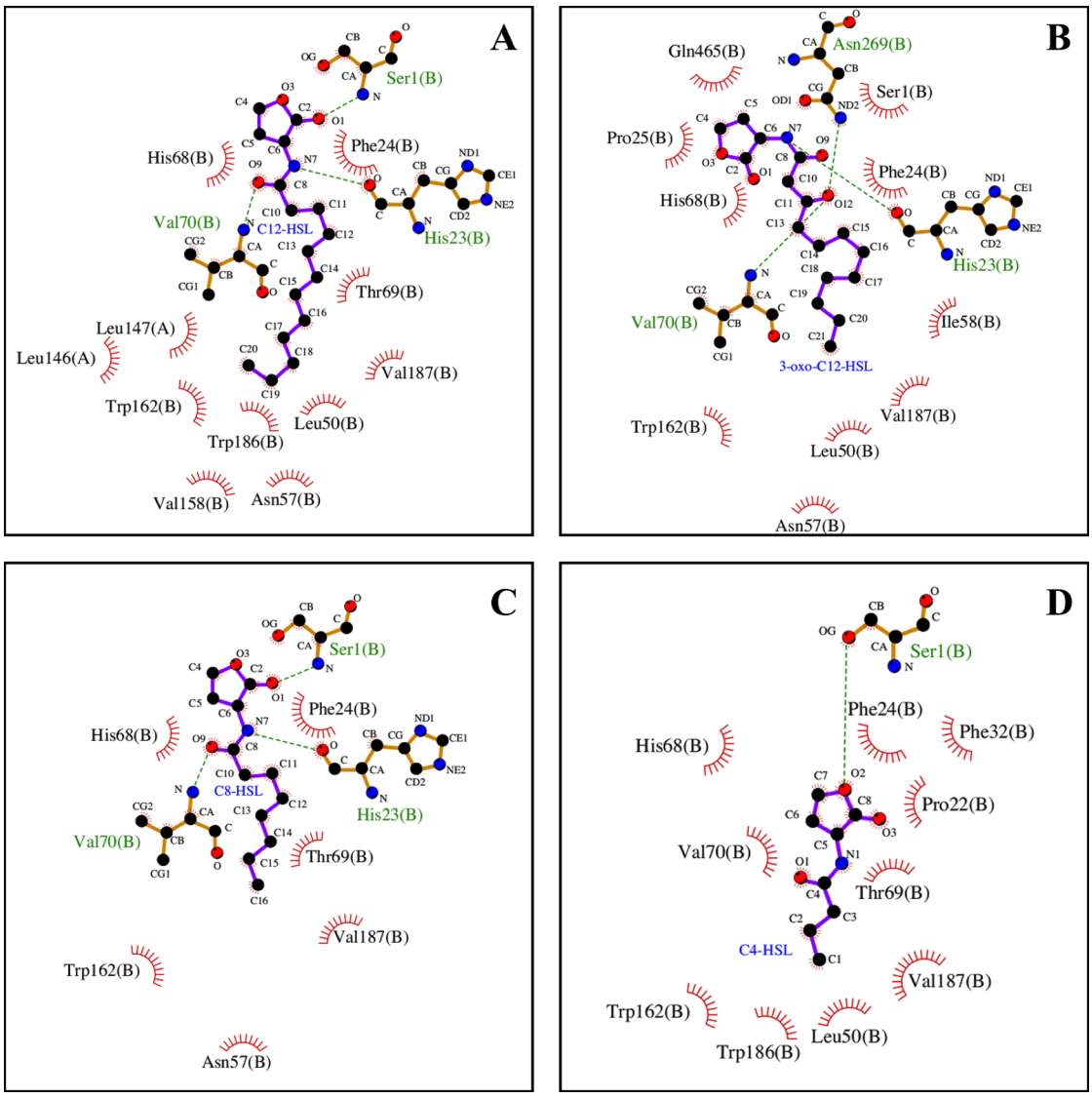

Figure 3.7 The ligand interactions of PvdQ with four different substrates from the GOLD docking. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex. The hydrogen bonds are represented by dashed lines, while the hydrophobic interactions are indicated by arcs with spokes.

3.3.4 Substrate binding modes and binding affinity analysis from Autodock Vina docking

Vina docking is a widely used open-source molecular docking method, which has the best scoring power.[165] The scoring power indicates the ability of a scoring function to rank the binding capabilities, which is calculated by Pearson correlation coefficient (r_p) and Spearman's rank correlation coefficient (r_s). Vina docking possesses an r_p/r_s of 0.564/0.580 for the top scored poses and r_p/r_s of 0.569/0.584 for best poses[165]. Since protein flexibility is very important to obtain high binding accuracy, several residues lining the binding pocket were treated as flexible in this docking study. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each ligand, the docking score, and the orientation of the binding poses were also used to select biologically reasonable binding poses. The data in **Table 3.6** indicate that the binding affinities of C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL are -8.3 kcal/mol, -7.3 kcal/mol, -6.3 kcal/mol, and -5.1 kcal/mol, respectively. The distances between the Oγ atoms of the catalytic center Serβ1 and the carbonyl carbon atoms of the amide bond of C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL are 3.2 Å, 3.4 Å, 7.3 Å, and 8.2 Å, respectively. These binding affinities are also positively correlated with the *N*-acyl chain lengths, which is consistent with the previous discussions.

Table 3.6 Distance between the catalytic center Serβ1 Oγ and the carbonyl carbon atom of the amide bond of each ligand, and docking scores obtained from the Vina docking

Vina docking results		
Complex Names	Distance (Å)	Docking score (kcal/mol)
C12-HSL	3.2	-8.3
3-oxo-C12-HSL	3.4	-7.3
C8-HSL	7.3	-6.3
C4-HSL	8.2	-5.1

The sliced hydrophobicity surfaces of PvdQ bound with different ligands are generated and represented in **Figure 3.8**. The tail groups of C12-HSL and 3-oxo-C12-HSL substrates are all located in the hydrophobic binding pockets. The head groups of C12-HSL and 3-oxo-C12-HSL substrates are all toward to the deep solvent-accessible clefts of the PvdQ protein. These orientations of C12-HSL and 3-oxo-C12-HSL substrates enable the carbonyl carbon atom of the amide bond be close to the catalytic center Serβ1, which increases the possibility for the catalysis to occur. The head group and tail group of the C8-HSL substrate are all buried inside the binding pocket. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of C8-HSL is far from hydrogen bonding distance. For the C4-HSL substrate, it is totally immersed inside the PvdQ protein, which is enclosed by the hydrophobic surface. The different binding patterns of C8-HSL and C4-HSL demonstrate the low catalytic efficiency of the enzyme with these substrates.

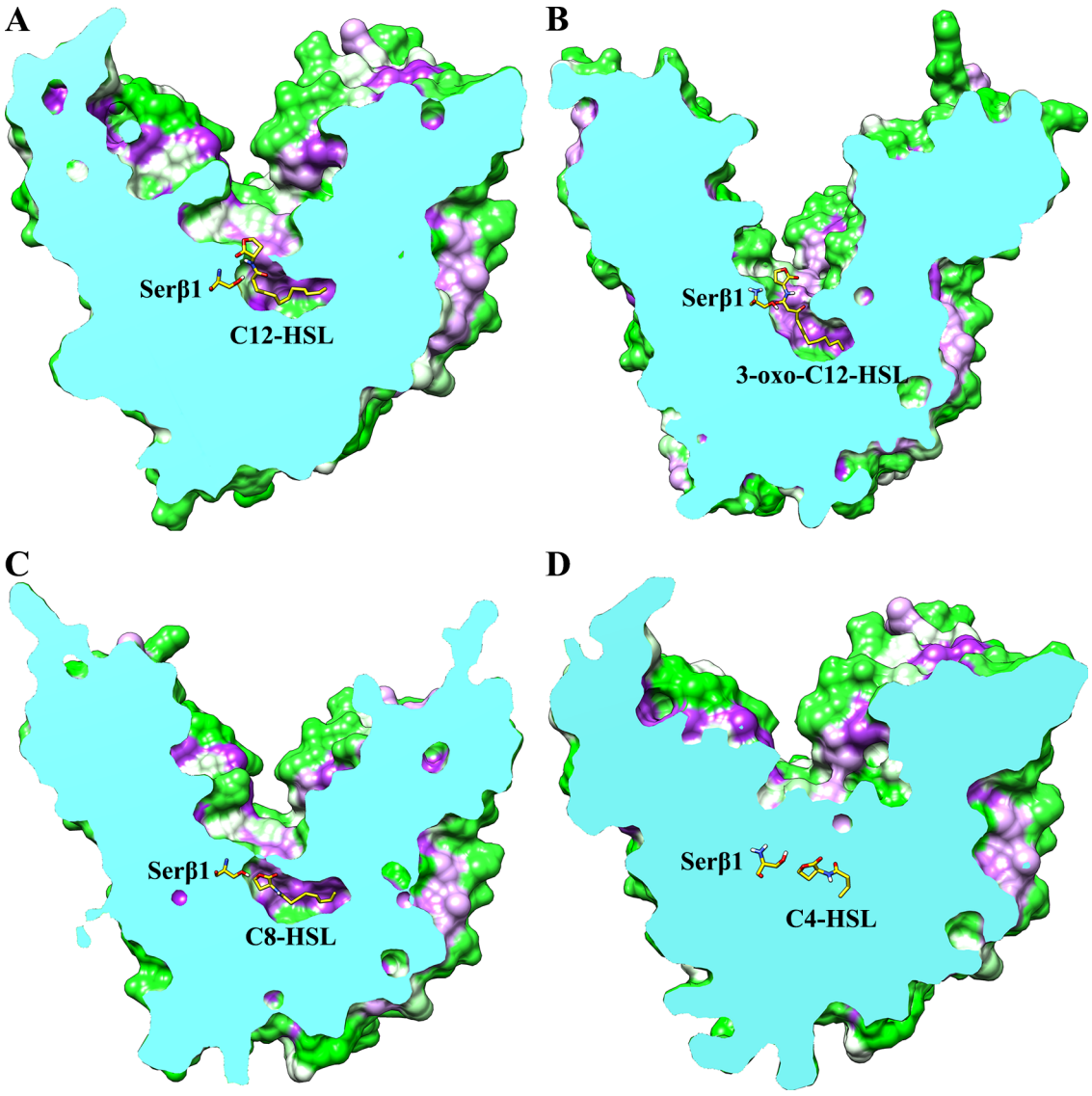

Figure 3.8 Sliced hydrophobicity surface of PvdQ binding with different substrates from the Vina docking. Green represents the most hydrophilic, while purple represents the most hydrophobic (color changes from green to white to purple). Serβ1 is the catalytic center. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex.

As shown in **Figure 3.9**, the position of each ligand after docking is buried inside the binding pocket of PvdQ. Slightly different side-chain movements for residues Serβ1, Valβ70, Trpβ186 and Leuα146 of each complex are observed, which can change the

volume of binding pocket to accommodate different substrates. The side-chains of Pheβ24 and Asnβ269 do not show substantial movement in the active site. The distances between the Oγ atoms of the catalytic center Serβ1 and the carbonyl carbon atoms of the amide bond of C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL are 3.2 Å, 3.4 Å, 7.3 Å, and 8.2 Å, respectively. These distances are negatively correlated with the *N*-acyl chain length. The Vina docking results illustrates that C12-HSL has the closest distance and the most favorable docking score. The 3-oxo-C12-HSL substrate is less energetically favorable as compared to the C12-HSL complex due to the extra 3-oxo substituent. The distances for the C8-HSL and C4-HSL poses are far beyond hydrogen bonding capability and have the least favorable docking scores, which is consistent with previous study.[205,207]

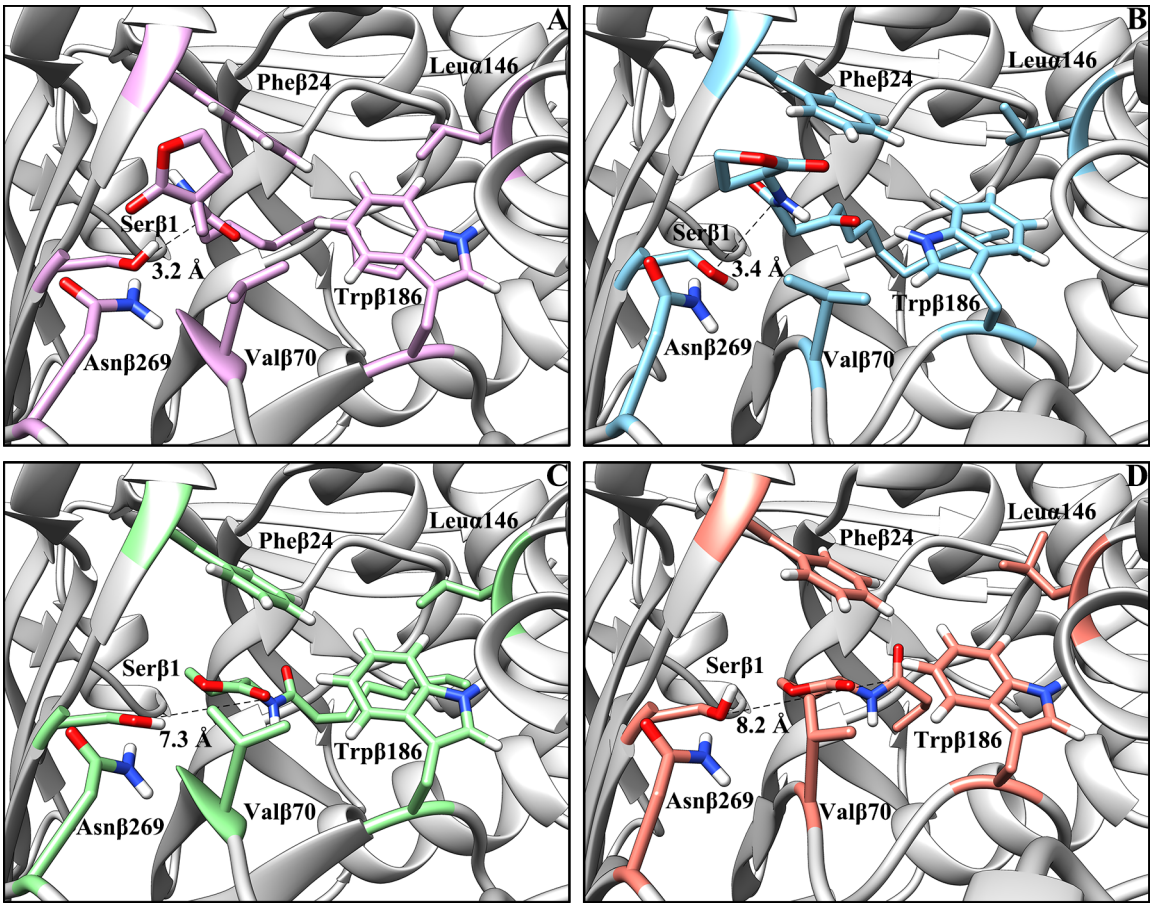

Figure 3.9 The docking poses of PvdQ with four different substrates from the Vina docking. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex. Residues treated flexible during docking are displayed in stick. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each ligand was measured.

The protein-ligand interaction diagrams are generated in **Figure 3.10**. There are three hydrogen bonds formed between the C12-HSL ligand and residues Serβ1, Hisβ23, and Asnβ269 of PvdQ. Residues Serβ1, Hisβ23, and Asnβ269 play important roles in stabilizing the head groups. Residues Hisβ23 and Trpβ186 form two hydrogen bonds with the 3-oxo-C12-HSL substrate. The hydrogen bond between residue Trpβ186 and the extra 3-oxo substituent of 3-oxo-C12-HSL help the substrate to bury deeper inside the binding pocket, which explains the relatively short distance between the Oγ atom of the

Serβ1 and the carbonyl carbon atom of the amide bond of 3-oxo-C12-HSL. There is only one hydrogen bond formed between C8-HSL and residue Serβ1. Conversely, the C4-HSL ligand is buried inside the binding pocket with zero hydrogen bond with the PvdQ protein. These results from Vina docking are also consistent with the experimental data that the PvdQ protein prefers different substrates with different acyl chains.[205,207]

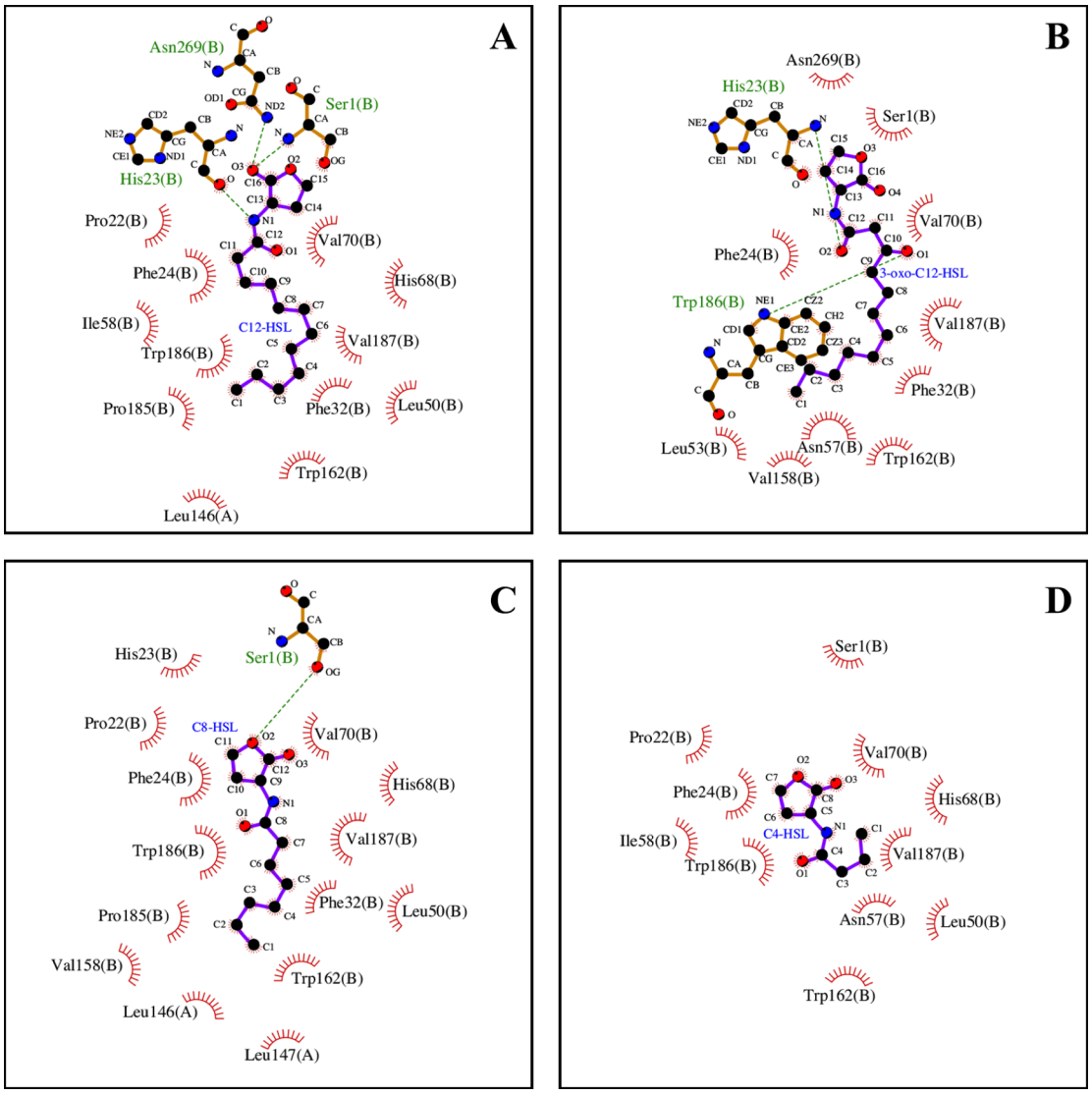

Figure 3.10 The ligand interactions of PvdQ with four different substrates from the Vina docking. A: PvdQ-C12-HSL complex B: PvdQ-3-oxo-C12-HSL complex C: PvdQ-C8-HSL complex D: PvdQ-C4-HSL complex. The hydrogen bonds are represented by dashed lines, while the hydrophobic interactions are indicated by arcs with spokes.

3.4 Conclusions

Molecular docking calculation is a widely used computational method in structure-based drug design. It can be applied to predict the binding mode and binding affinity of a substrate with the target protein. It plays an important role in studying the biological activities and correlations of protein and ligand structures. Therefore, three of the most popular docking programs are used to study the interactions of different ligands bound to the PvdQ protein. Even though the above experimental results of the three programs may seem redundant, it is very important to show the consistency of different docking programs. Otherwise, it is difficult to evaluate the docking performances

From the above discussions, we can see that results from three different docking programs are highly consistent with each other and previous studies,[205, 207] which means the docking results are reasonable and trustworthy. All three programs have shown that the binding affinity is positively correlated with the *N*-acyl chain length of each ligand. The positions of the head groups of C12-HSL and 3-oxo-C12-HSL are mainly toward the deep solvent-accessible clefts between the two knobs of PvdQ, while the tail groups are buried inside the hydrophobic binding pockets. The entire C8-HSL and C4-HSL substrates are completely positioned in the hydrophobic binding pockets. The movements of the side-chains of residues Serβ1, Hisβ23, Valβ70, Trpβ186, Asnβ269 and Leuα146 affect the volumes of the binding pockets. There are three hydrogen bonds formed between the C12-HSL ligand and residues Serβ1, Hisβ23, Valβ70, or Asnβ269 of the PvdQ protein, while fewer hydrogen bonds are formed with other ligands. Analyses of the results from three molecular docking calculations demonstrate that PvdQ has different

substrate binding specificities toward different ligands. The binding mode and binding affinity analysis indicate that C12-HSL has a higher probability to make contact with the catalytic center than other ligands, which explains its higher catalytic efficiency and thereby supporting the accuracy of the docking calculations. These docking results are consistent with the experimental data that the PvdQ protein prefers HSLs with long acyl chains and without the 3-oxo substituent. [205,207]

Chapter 4 Unraveling the substrate binding specificity of quorum-quenching acylase PvdQ

4.1 Introduction

P. aeruginosa and *B. cenocepacia* have drawn wide attention recently due to the seriousness of infection and high fatality rate when co-infected with other bacteria.[204,205] The enzymatic degradation of HSL signaling molecules used by various Gram-negative bacteria opens up new ways to fight against bacterial infections proven by previous study.[205] PvdQ is an AHL acylase with the ability to hydrolyze the amide bond of AHL signaling molecules.[207] However, the details of the binding interactions of PvdQ with bacterial substrates are not fully understood. Previous studies demonstrated that the PvdQ has different preferences for substrates or inhibitors with different acyl chain lengths and substituents (**Table 4.1**).[207,208] It prefers to hydrolyze the AHLs with a long acyl chain, such as 12 or 14 carbons in the tail group, rather than AHLs with a short tail group or a 3-oxo substituent, such as C8-HSL, C4-HSL, and 3-oxo-C12-HSL.

In Clevenger's study, they used the n-alkylboronic acid inhibitors to demonstrate that the C6-$B(OH)_2$ and C8-$B(OH)_2$ form tetrahedral adducts and bind to the acyl tail pocket, while the C2-$B(OH)_2$ and C4-$B(OH)_2$ form trigonal planar adducts and bind to the head binding pocket.[208] However, the structure of the n-alkylboronic acid inhibitors are structurally different from AHL ligands and the absence of a PvdQ crystal structure with the entire AHL ligand makes it necessary to use other methods to study the determinants

of the substrate binding specificity of PvdQ using native AHL ligands. Moreover, it is common knowledge that protein structures are not static. The dynamical behavior of proteins is essential for gaining a more complete understanding of its function.[214] Although the general catalytic mechanism of PvdQ has been revealed by structural studies, an understanding of the intrinsic molecular dynamical motions of PvdQ induced upon different ligand binding is currently lacking. The substrate binding specificity determinants of the quorum-quenching enzyme PvdQ with the different bacterial ligands are unknown and unintuitive.[202]

The current work addresses this by providing a better understanding of the dynamical characteristics of the PvdQ complex; thus aiding the future design of new inhibitors. In this study, multiple molecular docking calculations were performed to obtain favorable binding poses and binding affinities of PvdQ with different ligands (C12-HSL, 3-oxo-C12-HSL, C8-HSL, and C4-HSL).[215] Subsequently, 300 ns molecular dynamics (MD) simulations were carried out for each substrate-bound and substrate-free PvdQ. On the basis of the MD-simulated structures, an approach through the combined use of principal component analysis (PCA), free energy landscape (FEL), residue interactions network (RIN), WebPSN analysis, and binding free energy calculation was performed to gain insight into the substrate binding specificity of PvdQ. The MD simulation approach is used to investigate the substrate binding modes and molecular motions in these complexes. PCA and FEL are capable of analyzing low energy structures of each complex. RIN and WebPSN analyses can help us to better understand the residue interactions and communication pathways. While the binding free energy calculation can provide some information about the energy changes of PvdQ upon different ligand

binding. As a result, this combined method can be effective in providing information to more fully understand the mechanism of substrate binding specificity and structure-dynamic-function relationship of the PvdQ enzyme.

Table 4.1 The steady-state rate constants of selected substrates and the ΔG_{bind} of selected n-alkylboronic acid inhibitors from previous studies.

Ligands	k_{cat}/K_M ($M^{-1}s^{-1}$)[a]	n-alkylboronic acid inhibitors	ΔG_{bind} (kcal/mol)[b]
C12-HSL	2.2×10^5	C12-$B(OH)_2$	-13.2 ± 0.1
3-oxo-C12-HSL	2.3×10^3	3-oxo-C12-$B(OH)_2$	Not determined
C8-HSL	2.2×10^2	C8-$B(OH)_2$	-9.23 ± 0.02
C4-HSL	Not determined	C3-$B(OH)_2$	-3.1 ± 0.2

[a]The steady-state rate constants of selected substrates from previous studies. [b]The ΔG_{bind} of selected n-alkylboronic acid inhibitors from previous studies. The n-alkylboronic acid inhibitor is structurally homologous to the HSL with one more carbon in the acyl chain.

4.2 Computational methods

4.2.1 Molecular dynamics simulations

4.2.1.1 Energy minimization after docking calculations

In order to avoid steric clashes and to refine the structure of the complex after molecular docking, the energetically favorable docking poses were subjected to an energy minimization, before MD simulation, using Gromacs 4.6.7 with the GROMOS 54a7

force field.[183,216] The topology file for each ligand was generated with the PRODRG program[217] with no energy minimization in order to keep the original coordinates of the ligand. Then the partial charge of each ligand was refined by a Firefly quantum chemistry calculation with the 6-31G** split valence basis set and the B3LYP hybrid functional through geometry optimization and a subsequent energy calculation.[218] Finally, the topologies of ligands were refined according to the above result and previous study in order to generate the GROMOS-compatible charges and charge groups.[219,220] The Generalized Born implicit solvent model was used since it can lower system complexity and reduce the calculation time. Double precision, including conjugate gradient and steepest descent methods were used to minimize the system energy. Energy minimization was carried out until the system converged to the maximum atomic force less than 10 $kJ \cdot mol^{-1} \cdot nm^{-1}$. The minimized structure of each complex was used as the starting coordinate for MD simulations.

4.2.1.2 System set-up and energy minimization

MD simulations were performed on each substrate-bound and substrate-free PvdQ using the Gromacs 4.6.7 program with the GROMOS 96 54a7 united-atom force field to reduce computational burden.[183,216] The protein-ligand complex was solvated in a cubic simple point charge (SPC) water box. The distance between the furthest protein atom and the water box edge was 12 Å.[221] Four sodium ions were added to neutralize the system (**Figure 4.1**). Each system was subjected to a step-wise energy minimization starting with the steepest descent method to ensure that the system had no steric clashes. The protocol of the step-wise energy minimization involved four steps: 1). receptor-ligand complex was constrained while water and ions moved freely, 2). heavy atoms of the complex were

constrained while water, ions, and hydrogens moved freely, 3). atoms of the complex main chain were constrained while the rest of the system moved freely, and 4). the entire system was subjected to an unconstrained energy minimization [222]. The minimization was carried out until the maximum atomic forces acting on the system converged to less than 100 kJ · mol^{-1} · nm^{-1}.

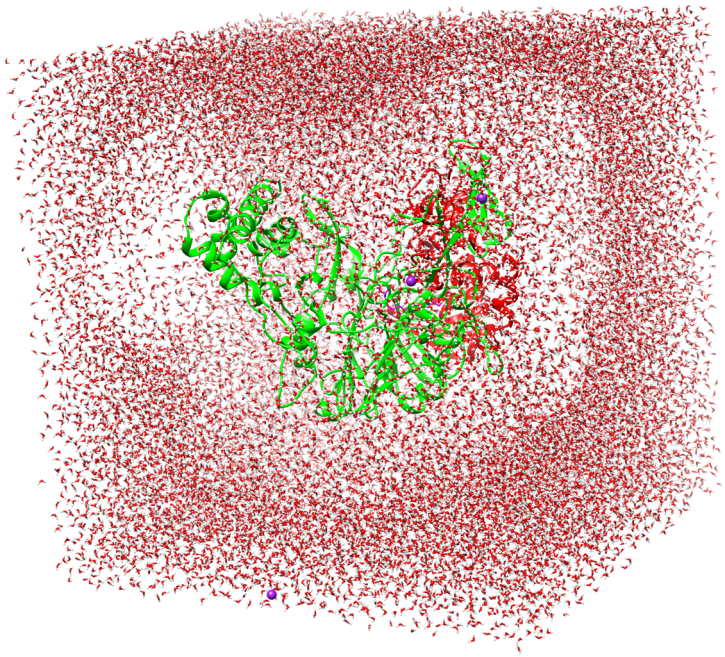

Figure 4.1 Representation of the simulation box of the substrate-bound PvdQ complex with water and sodium ions.

4.2.1.3 Equilibration and production simulation

After energy minimization, MD simulation of each system was performed under an isothermal-isochoric ensemble (constant Number of particles, Volume, and Temperature, NVT) for 100 ps to equilibrate the solvent and ions around the protein-ligand complex.

The systems were heated gradually from 0 K to 300 K using the single type simulated annealing method in the Gromacs program that is analogous to gradual heating methods used in other programs. Subsequently, the entire system was subjected to an isothermal-isobaric ensemble (constant Number of particles, Pressure, and Temperature, NPT) MD simulation for 200 ps to stabilize the pressure of the system. The harmonic position restraints on the heavy atoms were gradually released from 1000 $kJ \cdot mol^{-1} \cdot nm^{-2}$ to 0 $kJ \cdot mol^{-1} \cdot nm^{-2}$ to carefully relax the system. Finally, production runs of MD simulation were carried out under the NPT ensemble with no restraints on the entire system. Periodic boundary conditions were used to minimize edge effects. The particle mesh Ewald method was employed to handle the long-range electrostatics beyond the cutoff[223] and the LINCS algorithm was used to constrain all bond lengths.[224] The cutoff for short-range electrostatics was 0.9 nm and the cutoff for short-range van der Waals was 1.4 nm. The temperature was maintained at 300 K by the V-rescale method, while the pressure was kept at 1 bar by the Berendsen algorithm during the simulation.[182] The time step was set at 2 fs. The trajectories of each system were saved every 2 ps for later analysis. The entire production simulation was carried out for 300 ns (**Table 4.2**). The results were analyzed using Gromacs utilities.

Table 4.2 MD simulations of each substrate-bound PvdQ complex from three different docking

MD simulations from IFD poses				
Complex system	C12-HSL	3-oxo-C12-HSL	C8-HSL	C4-HSL
PvdQ residues	710	710	710	710
Ligand atoms	21	22	17	13
Water molecules	39689	39259	39885	39739
Total Na^+ atoms	4	4	4	4
Total atoms	126191	124902	126775	126333
EM	Until convergence	Until convergence	Until convergence	Until convergence
EQ NVT ensemble	100 ps	100 ps	100 ps	100 ps
EQ NPT ensemble	200 ps	200 ps	200 ps	200 ps
Production run	300 ns	300 ns	300 ns	300 ns
MD simulations from GOLD docking poses				
Complex system	C12-HSL	3-oxo-C12-HSL	C8-HSL	C4-HSL
PvdQ residues	710	710	710	710
Ligand atoms	21	22	17	13
Water molecules	39251	39922	39919	39430
Total Na^+ atoms	4	4	4	4
Total atoms	124877	126891	126877	125406
EM	Until convergence	Until convergence	Until convergence	Until convergence
EQ NVT ensemble	100 ps	100 ps	100 ps	100 ps
EQ NPT ensemble	200 ps	200 ps	200 ps	200 ps
Production run	300 ns	300 ns	300 ns	300 ns
MD simulations from Vina docking poses				
Complex system	C12-HSL	3-oxo-C12-HSL	C8-HSL	C4-HSL
PvdQ residues	710	710	710	710
Ligand atoms	21	22	17	13
Water molecules	39775	39718	39750	39930
Total Na^+ atoms	4	4	4	4
Total atoms	126449	126279	126370	126906
EM	Until convergence	Until convergence	Until convergence	Until convergence
EQ NVT ensemble	100 ps	100 ps	100 ps	100 ps
EQ NPT ensemble	200 ps	200 ps	200 ps	200 ps
Production run	300 ns	300 ns	300 ns	300 ns

4.2.2 Stability of the complexes

4.2.2.1 Root mean square deviation (RMSD)

RMSD of the backbone was calculated for the PvdQ-ligand complexes through the Gromacs analysis toolkit, g_rms.

4.2.2.2 Root mean square fluctuation (RMSF)

RMSF of the alpha carbons were measured with the Gromacs tool g_rmsf during the 300 ns production runs for each complex.

4.2.2.3 Radius of gyration (Rg)

Rg is the root mean square distance of the atomic positions relative to the center of the mass of the protein. The Rg of each complex was calculated with g_gyrate to examine the compactness of each complex during the entire simulation.

4.2.3 Hydrogen bond occupancy

The hydrogen bonds of each substrate bound PvdQ complex were calculated using g_hbond module within GROMACS program. The hydrogen bonds occupancy was calculated by a python code named readHBmap.py. A hydrogen bond was assigned when the distance between donor and acceptor was less than 3.5 Å as well as the angle of donor-hydrogen-acceptor was more than 140°.[225]

4.2.4 Distance analysis

The distance between the Oγ atom of catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each ligand during the 300 ns MD simulations were calculated by g_dist module within GROMACS program.

4.2.5 Buried surface area

To observe the surface interactions between PvdQ and different ligands, the buried surface areas of different ligand complexes were calculated using the g_sas module within the GROMACS program.

4.2.6 Principal component analysis (PCA) and free energy landscape (FEL)

The PCA and FEL methods were used to analyze the minimal energy structure of each complex. The covariance matrix from alpha carbon fluctuation was calculated after elimination of the translational and rotational degrees of freedom. After diagonalization of the covariance matrix, the eigenvalues and eigenvectors were obtained. This essential dynamics analysis was performed with Gromacs g_covar and g_anaeig commands for each 300 ns MD simulation. The first few eigenvectors, which correspond to the highest eigenvalues, are called the principle components (PC) since they represent the correlated collective motions of the protein-ligand system. The projections of the first two principal components, which account for about 40%, revealed overall correlated motions in each complex. In order to obtain the energy minimum structures during the 300 ns MD simulations, the free energy landscape along the first two principal components was generated by the g_sham module within the GROMACS program[226] and plotted with MATLAB.

4.2.7 Protein dynamic fingerprint

In order to investigate the dynamical motions of PvdQ induced upon different ligand binding, the protein dynamic fingerprints were calculated with the g_covar and g_anaeig modules within the GROMACS program and plotted with Visual Molecular Dynamics

program. The first principal component was used to analyze the protein dynamic fingerprint since it collectively accounts for around 30% of all dynamical motions. The direction of the arrow indicates the direction of the motion, while the length of the arrow represents the strength of the movement.

4.2.8 Residue interaction network

Furthermore, in order to understand the interactions between the residues lining the binding pocket, the residue interaction network (RIN) was produced using the Residue Interaction Network Generator (RING).[189] The RING-2.0 program is able to identify six types of interactions, namely, hydrogen-bonds, van der Waals interactions, disulfide bridges, salt bridges, π-π stacking, and π-cation interactions. The output files were visualized with the Cytoscape program.[227] In the interaction network, each node is represented as a residue or a ligand molecule, while each edge line between two nodes is denoted as a noncovalent interaction between two residues. The different types of noncovalent interactions are indicated as different colored edges. In order to further analyze the topological parameters of each RIN, the closeness centrality and betweenness centrality were calculated with the NetworkAnalyzer plugin of the Cytoscape program.[228] Closeness centrality indicates how fast information spreads from one node to other reachable nodes in the interaction network.[229] The betweenness centrality describes the amount of control one node exerts over the other nodes in the interaction network.[230] In an interaction network, the values of closeness centrality were mapped to the color of the nodes, mapping low values to green and high values to red. The values of betweenness centrality were mapped to the size of the nodes, mapping low values to small sizes and

high values to large sizes.

4.2.9 WebPSN analysis

As an extension of the residue interaction network analysis implemented in the RING program, we used WebPSN to represent the networks on interacting residues in system to investigate the potential role of structurally important residues, the protein stability, and communication.[190,193,231] This method begins with a protein structure graph; then searches for all possible communication paths and filters the results according to the cross-correlation of atomic motions to obtain shortest communication pathways that traverse the protein structure. Meta paths include the most recurrent nodes and links in the full communication pathways, which is important for acquiring the global picture of the structural communication of the whole system. The representative structure of the most populated cluster of each complex was used for the generation of the residue communication pathway. The node and link recurrence cutoff used in building the meta paths was ≥ 30%. Each spherical node was centered on the Cα-atom of one residue, whose diameter is proportional to the number of links generated by the node. The link thickness is also proportional to link frequencies in the communication paths.

4.2.10 Binding free energy analysis

In order to better understand the energy changes of PvdQ induced upon binding of different ligands, the Molecular Mechanics Poisson-Boltzmann (or Generalized Born) Surface Area (MM-PBSA or MM-GBSA) method can be used to calculate the individual energy components and total binding free energy. Both MM-PBSA and MM-GBSA are widely used approaches in free energy calculations. A thorough comparison of MM-

PBSA and MM-GBSA methods had also been conducted where the overall prediction accuracies of the two methods had been comprehensively investigated along with the studies of the entropy effects, improved docking performance with high solute dielectric constant (i.e., ε =2, or 4), and the predicting of binding free energies for protein-protein and protein-RNA systems.[232-236] In this study, in order to compare the free energy changes of substrate-bound PvdQ, the binding free energy for each complex was computed using the g_mmpbsa package in the Gromacs utility.[237] The binding free energy can be decomposed into different parts, containing van der Waals, electrostatic, polar, and nonpolar free energy contributions. The entropy contribution at 300K (-TΔS) was calculated by normal mode analysis from energy-minimized structures generated in the MD simulations.[238-240] The free energy equation is shown below. In the equation, γ was 0.00542 kcal/(mol·Å^2), while β was 0.92 kcal/mol. The bondi set of atomic radii was used in this study.[237] The dielectric constant values for the solute and solvent were set to be 2 and 80, respectively. A total of 3000 snapshots were extracted from the 300 ns production trajectories at intervals of 0.1 ns and were used for the MM-PBSA calculations.

$$\Delta G_{binding} = G_{complex} - (G_{receptor} + G_{ligand}) \qquad \text{(Equation 4.1)}$$

$$\Delta G_{binding} = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \qquad \text{(Equation 4.2)}$$

$$\Delta E_{MM} = \Delta E_{vdW} + \Delta E_{ele} \qquad \text{(Equation 4.3)}$$

$$\Delta G_{sol} = \Delta G_{pol} + \Delta G_{nonpol} \quad \text{(Equation 4.4)}$$

$$\Delta G_{nonpol} = \gamma(SASA) + \beta \quad \text{(Equation 4.5)}$$

$$\Delta S = S_{complex} - (S_{receptor} + S_{ligand}) \quad \text{(Equation 4.6)}$$

4.3 Results and discussion

MD simulations are a critical method to investigate the global motions and dynamical behaviors of protein-ligand complexes. The favorable docking modes obtained from multiple molecular docking calculations were used as the starting structures for MD simulations. The complexes were allowed to move and form more favorable conformations during MD simulations to improve the interactions between protein and ligand. Therefore, MD simulations were carried out to analyze the specific interactions and overall molecular motions of PvdQ induced upon different ligand binding. The average structural properties from the 300 ns MD simulation results derived from three different docking programs are listed in **Table 4.3**. The results of MD simulations from three different docking programs were consistent with each other and the Schrödinger program is one of the most robust programs on pose predictions with a nearly 90% rate (SRtsp/SRbp, where SRtsp and SRbp are the success rates for the top scored poses and best poses)[165] and performs best when binding sites are mainly influenced by hydrophobic contacts.[167] Hereafter, the MD simulation results from the Schrödinger IFD poses were selected for further analysis and discussion.

Table 4.3 Average structural properties of different substrate-bound PvdQ during 300ns MD simulations from three different docking

MD results from Schrödinger IFD poses				
Complex	C12-HSL	3-oxo-C12-HSL	C8-HSL	C4-HSL
RMSD (Å)	2.15±0.28	2.40 ± 0.28	2.29± 0.29	2.74± 0.43
RMSF (Å)	1.17±0.77	1.25± 0.84	1.22± 0.78	1.32± 0.88
Rg (Å)	26.18±0.12	25.96±0.19	25.91±0.17	25.99±0.15
Distance (Å)	3.40± 0.35	4.58± 1.13	7.91± 0.80	9.81± 1.26
Buried surface area (Å^2)	430.28±14.13	409.49±14.28	324.56±11.20	246.43±9.79
FEL (kJ/mol)	-8.55	-8.43	-8.02	-8.17
ΔG_{bind} (kcal/mol)	-12.74±2.17	-10.84±2.50	-9.63±2.11	-5.34±2.33
MD results from GOLD poses				
Complex	C12-HSL	3-oxo-C12-HSL	C8-HSL	C4-HSL
RMSD (Å)	2.15± 0.37	2.16± 0.36	2.41 ± 0.39	2.76± 0.50
RMSF (Å)	1.27±0.72	1.21±0.83	1.28± 0.87	1.42± 0.96
Rg (Å)	26.15±0.14	26.29±0.13	26.06±0.20	26.14±0.14
Distance (Å)	3.36±0.41	4.26±0.76	3.37± 0.28	13.32±2.45
Buried surface area (Å^2)	425.17±14.64	408.72±14.14	331.48±11.30	283.02±9.52
FEL (kJ/mol)	-8.66	-7.16	-7.87	-7.54
ΔG_{bind} (kcal/mol)	-11.54±2.75	-10.66±2.76	-8.94±2.06	-6.42±2.31
MD result from Vina poses				
Complex	C12-HSL	3-oxo-C12-HSL	C8-HSL	C4-HSL
RMSD (Å)	2.20± 0.26	2.13± 0.26	2.21± 0.23	2.24 ± 0.42
RMSF (Å)	1.26Å±0.61	1.23± 0.83	1.28± 0.92	1.30± 0.96
Rg (Å)	26.11±0.18	26.04±0.20	26.05±0.14	25.11±0.15
Distance (Å)	3.36± 0.40	4.47± 0.61	5.60± 1.23	9.49± 0.60
Buried surface area (Å^2)	424.51±13.52	416.18±14.88	311.48±12.31	280.06±8.90
FEL (kJ/mol)	-8.77	-8.17	-8.43	-8.55
ΔG_{bind} (kcal/mol)	-10.40±2.61	-8.65±2.81	-7.79±2.76	-4.50±2.01

4.3.1 The stability analysis for each system

Root mean square deviation (RMSD) is commonly used to measure the deviation of the system with respect to the starting structure over time. As illustrated in **Figure 4.2A**, the RMSD values from Schrödinger IFD poses deviate a small extent from their starting structures during the 300 ns production run. These results illustrate that each system reached equilibrium and retaine stable fluctuations during the simulations. The overall variation of the RMSD is not remarkable, although slight RMSD changes of each system are observed. The RMSD value for the C12-HSL complex is slightly smaller as compared to the values of 3-oxo-C12-HSL, C8-HSL, and C4-HSL complexes and substrate-free PvdQ. Therefore, the other substrate-protein complexes have more backbone deviation as compared to the C12-HSL complex and undergo conformational shifts at varying time points that cause relatively high trends in RMSD. Furthermore, the RMSD values of active site residues lining the binding pocket and the two knobs were calculated (**Figure 4.3**). The results exhibit that the binding of C12-HSL reduces the deviations of the active site and the two knobs. These results show that the conformational diversifications of different substrate-protein complexes are a direct result of deviations in both internal stability and structural dynamics induced upon different ligand binding.

To further investigate the individual residue flexibility of each substrate-protein complex, the root mean square fluctuation (RMSF) of the alpha carbons were measured during the 300 ns production runs for each complex (**Figure 4.2B**). From the RMSF plot, it can be seen that the overall fluctuations of the C12-HSL complex are slightly lower as compared to the other substrate-protein complexes and unliganded PvdQ. It is noticed that there are some dramatic fluctuations in the terminal regions (6-7, 166-169) due to the flexibility of

the unrestrained termini. In the PvdQ-C4 and PvdQ-C8 complexes, the large fluctuations in the surface loop regions at residues 111-126 (residues 280-285 in **Figure 4.2B**), 131-143 (residues 300-312 in **Figure 4.2B**), 212-219 (residues 381-388 in **Figure 4.2B**), 340-345 (residues 509-514 in **Figure 4.2B**), 390-398 (residues 559-567 in **Figure 4.2B**), and 428-434 (residues 597-603 in **Figure 4.2B**) of the PvdQ B chain may affect the stability of the complex. The binding of C4-HSL produces the highest fluctuations than the others and these large fluctuations could induce ligand instability, and in turn affect the stability of the PvdQ complex. Therefore, these results reveal that the binding of different ligands influence the fluctuations of the loops located in the two knobs. The movements of those loops are likely to affect the conformation of the active site and influence the dynamic stability of the entire PvdQ structure.

To further examine the effects of different ligand binding on PvdQ protein compactness, the radius of gyration (Rg) was computed of each complex (**Figure 4.2C**). Rg is the root mean square distance of the atomic positions relative to the center of the mass of the protein. The Rg of each complex was calculated to examine the compactness of each complex during the entire simulation. The respective mean Rg value for C12-HSL, 3-oxo-C12-HSL, C8-HSL and C4-HSL complexes are 26.2 ± 0.1 Å, 26.0 ± 0.2 Å, 25.9 ± 0.2 Å, and 26.0 ± 0.2 Å, which exhibit comparable values to the substrate-free PvdQ with a value of 26.0 ± 0.3 Å. These results demonstrate that the structures for the systems studied are compact and well folded, and that there is no meaningful change in Rg on ligand binding.

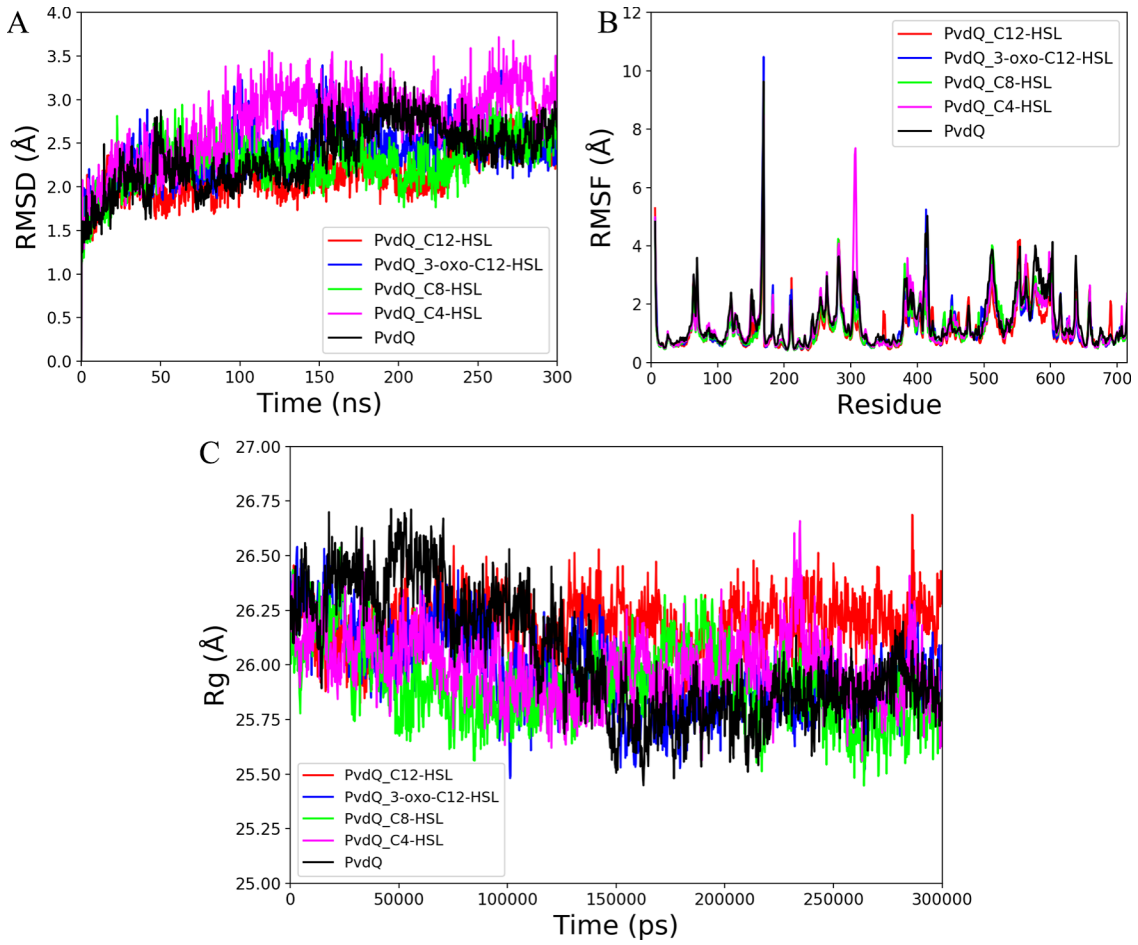

Figure 4.2 (A) RMSD, (B) RMSF and (C) Rg of substrate-bound and substrate-free PvdQ.

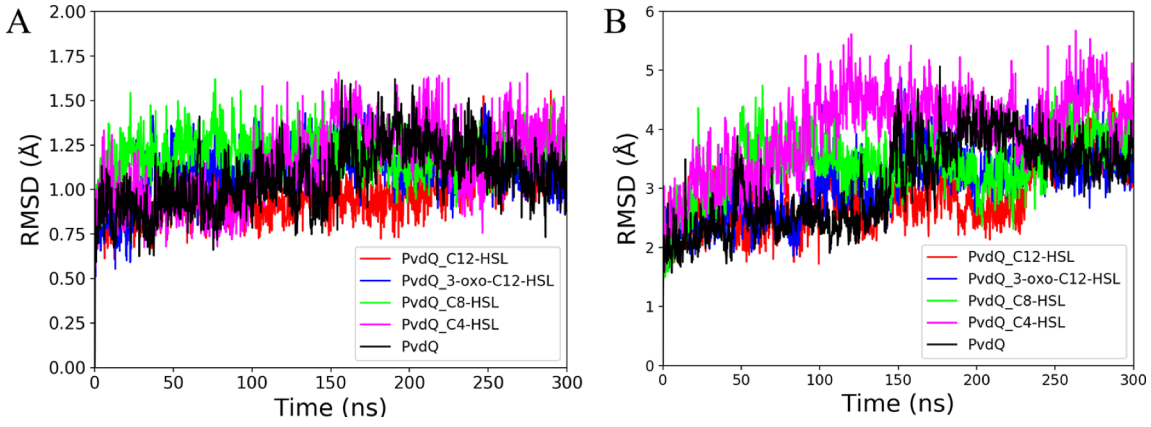

Figure 4.3 (A) RMSD of active site residues lining the binding pocket. (B) RMSD of the two knobs.

4.3.2 Hydrogen bond occupancies

In order to obtain additional insight into the protein-ligand interaction, hydrogen bond occupancies were calculated. These are a time ratio of the hydrogen bond existence during the entire simulation, which is a widely used method to indicate the frequency of the hydrogen bond interactions between the protein and ligand.[241] There are plenty of hydrogen bonds formed during the 300 ns simulations for each complex. The occupancies of hydrogen bond pairs greater than 5% are shown in **Figure 4.4**. The most populated hydrogen bond pairs formed for the C12-HSL complex are HSL(H1) – Hisβ23(O), Serβ1(HG) – HSL(O1), Serβ1(H1) – HSL(O1), and Valβ70(H) –HSL(O3), with occupancies of 86%, 74%, 49% and 22%, respectively. These hydrogen bonding interactions are stable and the residues involved in the bonding play an important role in ligand binding. The most populated hydrogen bond pairs formed for the 3-oxo-C12-HSL complex are HSL(H1) – Hisβ23(O), Asnβ269(D21) – HSL(O1), Serβ1(H1) – HSL(O1), Serβ1(HG) – HSL(O4), Serβ1(HG) – HSL(O1), Hisβ23(H) –HSL(O4), and Serβ1(H1) – HSL(O4), whose occupancies are 80%, 49%, 48%, 41%, 16%, 9%, and 8%, respectively. The 3-oxo-C12-HSL ligand has some additional hydrogen bonds due to the extra 3-oxo substituent – Serβ1 and Hisβ23. C8-HSL and C4-HSL form different hydrogen bonding patterns since the position to which they bind are far from the catalytic center. C4-HSL forms a hydrogen bond with Asnβ57 during the majority of the time in the simulation, which permits C4-HSL to diffuse around this position. These differences in the hydrogen bond occupancies could be useful in explaining the substrate specificity and helpful for further inhibitor design.

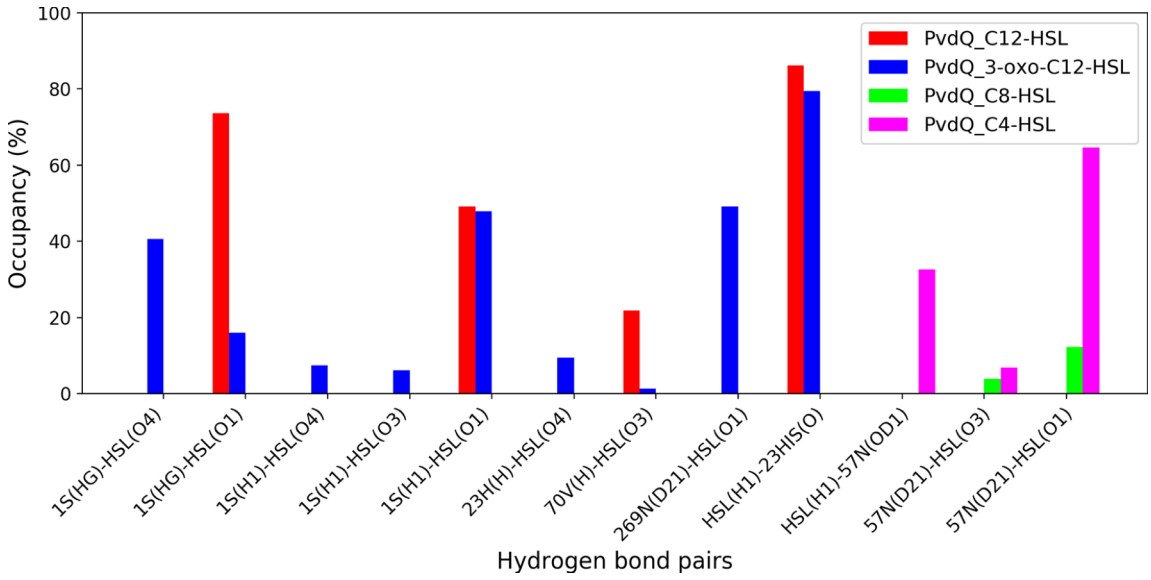

Figure 4.4 Hydrogen bond occupancy for four different ligands.

4.3.3 Distance analysis

Furthermore, the distance between the Oγ atom of catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of each ligand during the 300 ns MD simulations were calculated to analyze the contacts and catalytic potential between PvdQ and the various ligands (**Figure 4.5**).

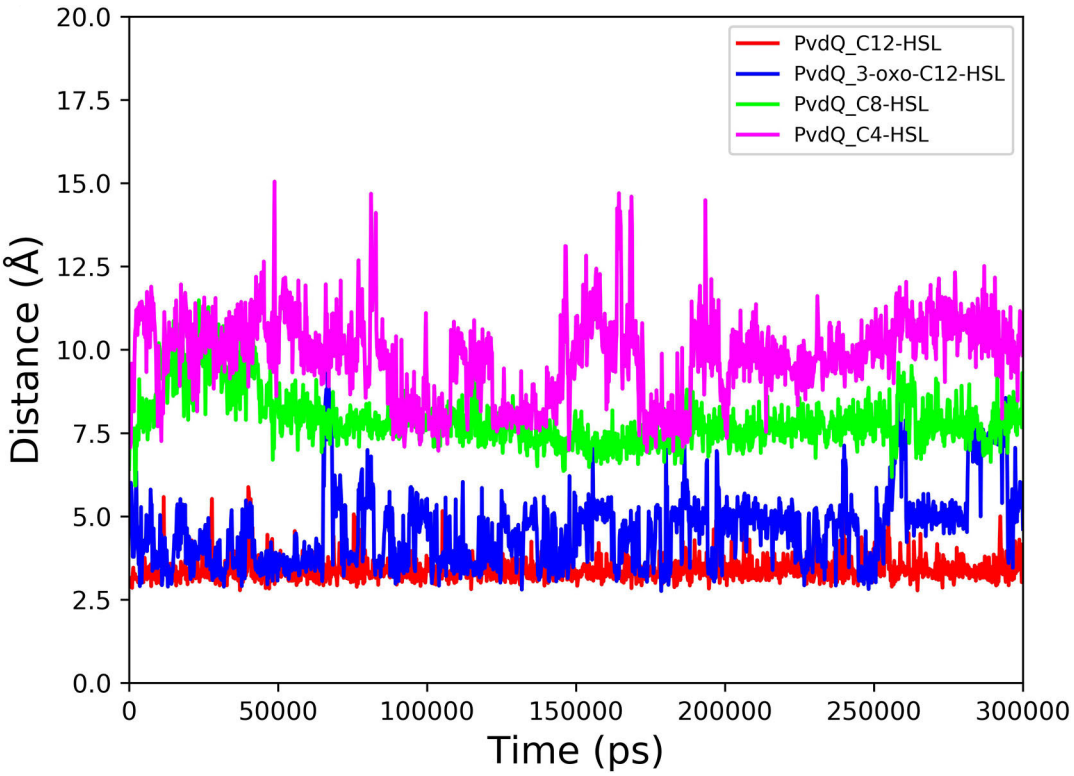

Figure 4.5 The distances between the Oγ atom of catalytic center Serβ1 and the carbonyl carbon atom of the amide bond.

The average distance for C12-HSL binding is 3.40 ± 0.35 Å during the 300 ns MD simulation, which is close to the catalytic center Serβ1 and has a higher chance to contact with it. **Figure 4.6** shows the occurrence probability for different distances of each complex during the 300 ns MD simulations. For the C12-HSL complex, 95.5% of the measured distances during the simulations are less than 4 Å, which is consistent with the catalytic efficiency of PvdQ for this ligand. The mean corresponding distances for 3-oxo-C12-HSL, C8-HSL and C4-HSL are 4.58 ± 1.13 Å, 7.91 ± 0.80 Å and 9.81 ± 1.26 Å, which are beyond hydrogen bonding distance, showing they have a lower probability to make contact with the catalytic center and should have lower catalytic efficiencies. This is consistent with the observation that PvdQ has different catalytic efficiencies toward

different ligands and has highest efficiency with C12-HSL.

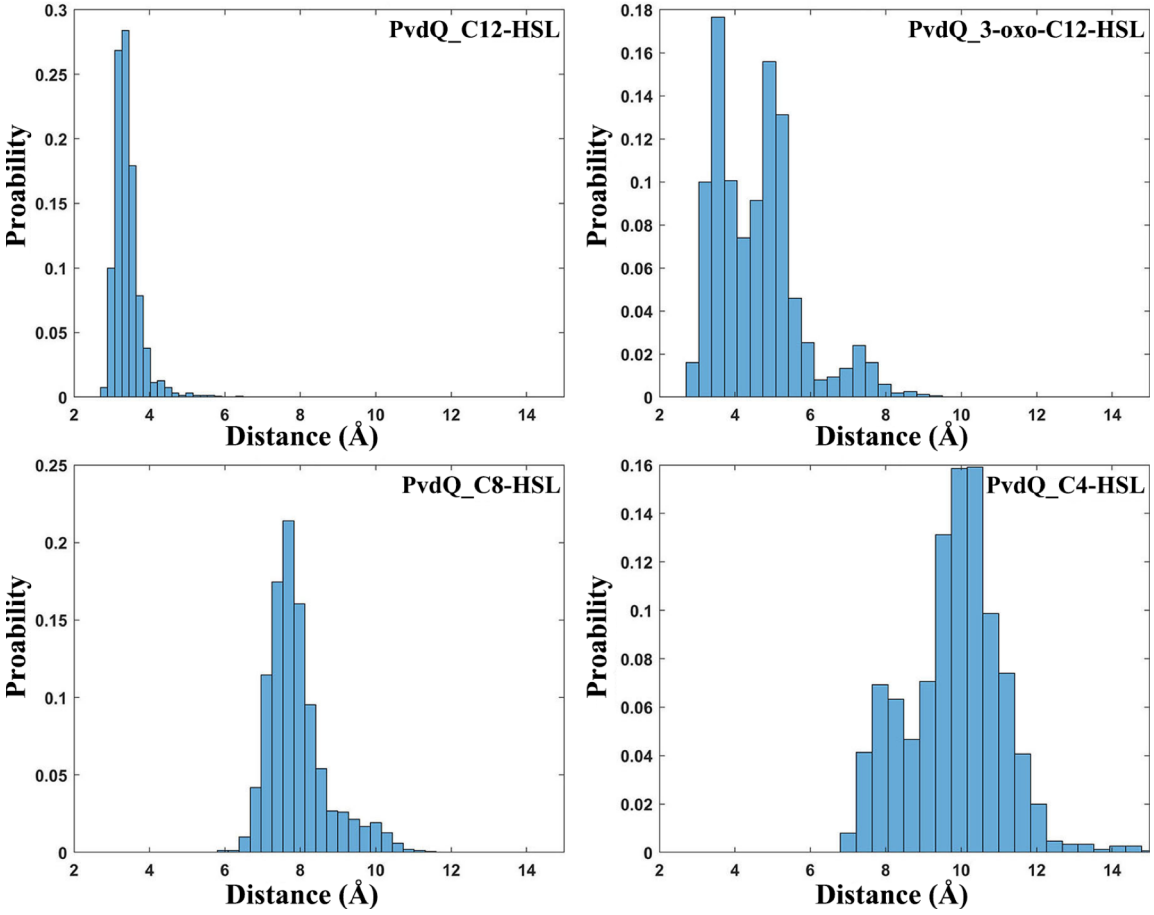

Figure 4.6 The occurrence probability for different distances during 300 ns MD simulations.

4.3.4 Buried surface area

To observe the surface interactions between PvdQ and different ligands, the buried surface areas of different ligand complexes were calculated for each complex (**Figure 4.7**). C12-HSL exhibits the largest total buried surface area with a mean value of 424.5 $Å^2$, which includes the largest hydrophobic buried surface area of the ligand among all

ligands. This result demonstrates that the entire acyl tail group of C12-HSL is buried inside the hydrophobic binding pocket and the hydrophobic interactions are maintained between C12-HSL and the PvdQ protein during the entire simulation. 3-oxo-C12-HSL binding reveals the largest hydrophilic buried surface area with a value of 110.0 $Å^2$, which indicates that the head group and the extra 3-oxo substituent interact with the polar surface area of the PvdQ protein. C8-HSL and C4-HSL have a shorter tail group, which is consistent with their buried surface areas.

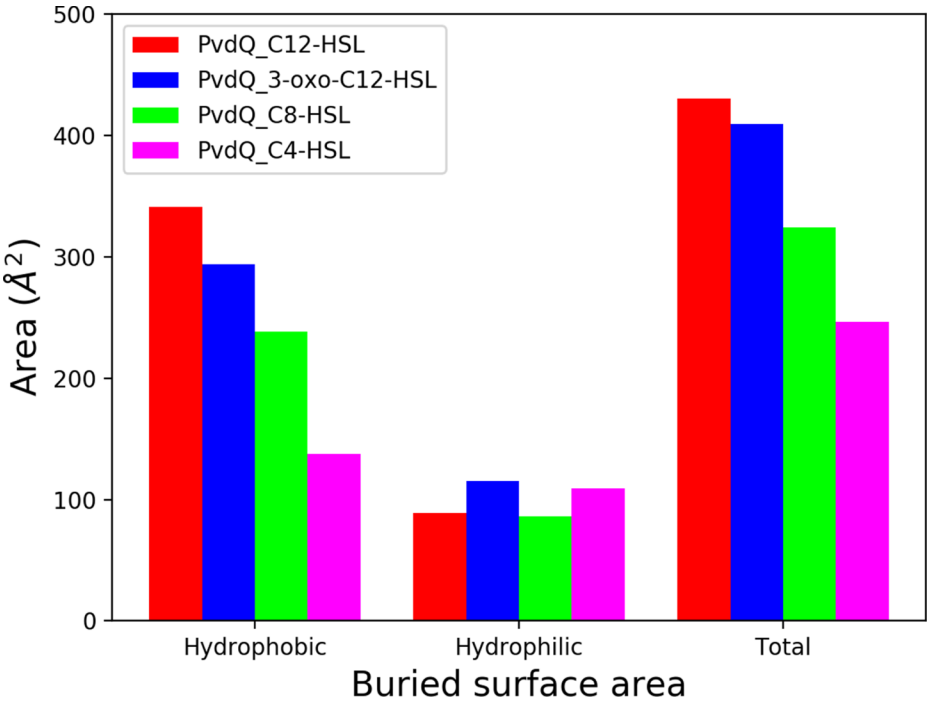

Figure 4.7 The buried surface area of different ligands.

4.3.5 Principal component analysis (PCA) and free energy landscape (FEL)

The PCA method is a widely used strategy to investigate dominant motions that lead to the global correlated dynamics of protein, induced upon different ligand binding. The projections of PC1 and PC2 were calculated to show the conformational distribution. The

FEL along the first two principal components was generated to investigate the energetics of the different conformations of the complexes sampled during the MD simulation. As shown in **Figure 4.8**, the direction of the motions and the probability of visited conformational regions of the complexes differ from each other. The blue region in all figures have the lowest Gibbs free energy. The size and shape of FEL vary among the different substrate-bound complexes, which indicates the stability differences of each complex. The more centralized, deeper, and smaller the blue areas of the "valley", the more stable is the complex.[242,243]

As shown in **Figure 4.8**, the complex of C12-HSL exhibits one small dominated energy minimum, which indicates that this complex is stabilized in a minimum energy conformational region. 3-oxo-C12-HSL, C8-HSL complexes, and PvdQ protein alone (**Figure 4.9**) mainly visit two energy minima, which demonstrates that these two complexes cross two subspaces and transform between distinct conformational states. For the C4-HSL complex, one dominated energy minimum spread out over a large portion of the free energy space is detected, which indicates that this complex had large conformational changes in the surrounding regions. Even though the depth of the "valley" of each system is similar to each other, the energy minimum value is positively correlated with its stability. The C12-HSL complex has a slightly deeper energy minimum with a value of -8.55 kJ/mol as compared to 3-oxo-C12-HSL, C8-HSL, and C4-HSL with values of -8.43 kJ/mol, -8.02 kJ/mol, and -8.17 kJ/mol, respectively. Therefore, these analyses illustrate that the binding of the C12-HSL compound results in a higher stability, which is in accord with previous experimental results.[207]

The corresponding minimum energy structure of each complex extracted from the MD

simulation trajectories is indicated with arrows. For the C12-HSL complex, there is one minimum energy structure. The C12-HSL ligand is tightly bound to the substrate binding pocket of the PvdQ protein. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of C12-HSL is 3.1 Å. The tail group forms strong hydrophobic interactions with the surrounding residues in the binding pocket. The head group forms hydrogen bonds with the side chain of Serβ1 and backbone of Hisβ23 and Valβ70. For the 3-oxo-C12-HSL complex, there are two minimum energy structures, which display a close resemblance to one another except for some fluctuations in the two knobs. Pheβ24 acts as a gate separating the solvent and the hydrophobic binding pocket. For the deeper minimum energy structure, there are more hydrogen bonds formed due to the "open" state of Pheβ24. The additional 3-oxo substituent is a polar group, which forms a hydrogen bond with the side chain of Serβ1. However, this extra 3-oxo substituent blocks the movement of the ligand. Therefore, the tail of 3-oxo-C12-HSL is not as deeply buried as C12-HSL inside of the hydrophobic binding pocket. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of 3-oxo-C12-HSL is 5.1 Å. For the C8-HSL complex, there are also two minimum energy structures, which are similar to each other except for the rearrangement of the substrate binding site. The entire structures of the C8-HSL and C4-HSL ligands are completely buried inside the binding pocket. The polar head group is surrounded by unfavorable hydrophobic residues and the shorter tail group contributes to the lower hydrophobic energy. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of C8-HSL and C4-HSL are 7.3 Å and 8.9 Å, respectively, which are far beyond hydrogen bonding distance. Previous

studies demonstrated that the n-alkylboronic acid inhibitors of PvdQ, C6-$B(OH)_2$ and C8-$B(OH)_2$, can form tetrahedral adducts and bind to the acyl tail pocket, which was more favorable than the shorter ligands. On the other hand, the C2-$B(OH)_2$ and C4-$B(OH)_2$ inhibitors can form trigonal planar adducts and bind to the catalytic pocket instead of the acyl-group binding pocket, which supported the observation that PvdQ prefers substrates with longer chains over short ones.[208] However, the results from this study show that the C4-HSL ligand is located inside the pocket and is stabilized by forming a hydrogen bond with the side chain of Asnβ57. The C4-HSL ligand has a short tail group, which is flexible, moving around inside the hydrophobic binding pocket. The difference between this result and the previous study may be due to the structural differences of the substrates. The n-alkylboronic acid inhibitors dose not have the head ring group nor the amide bond as compared to the native ligands studied herein. Therefore, the results from this study are compatible with the previous study.

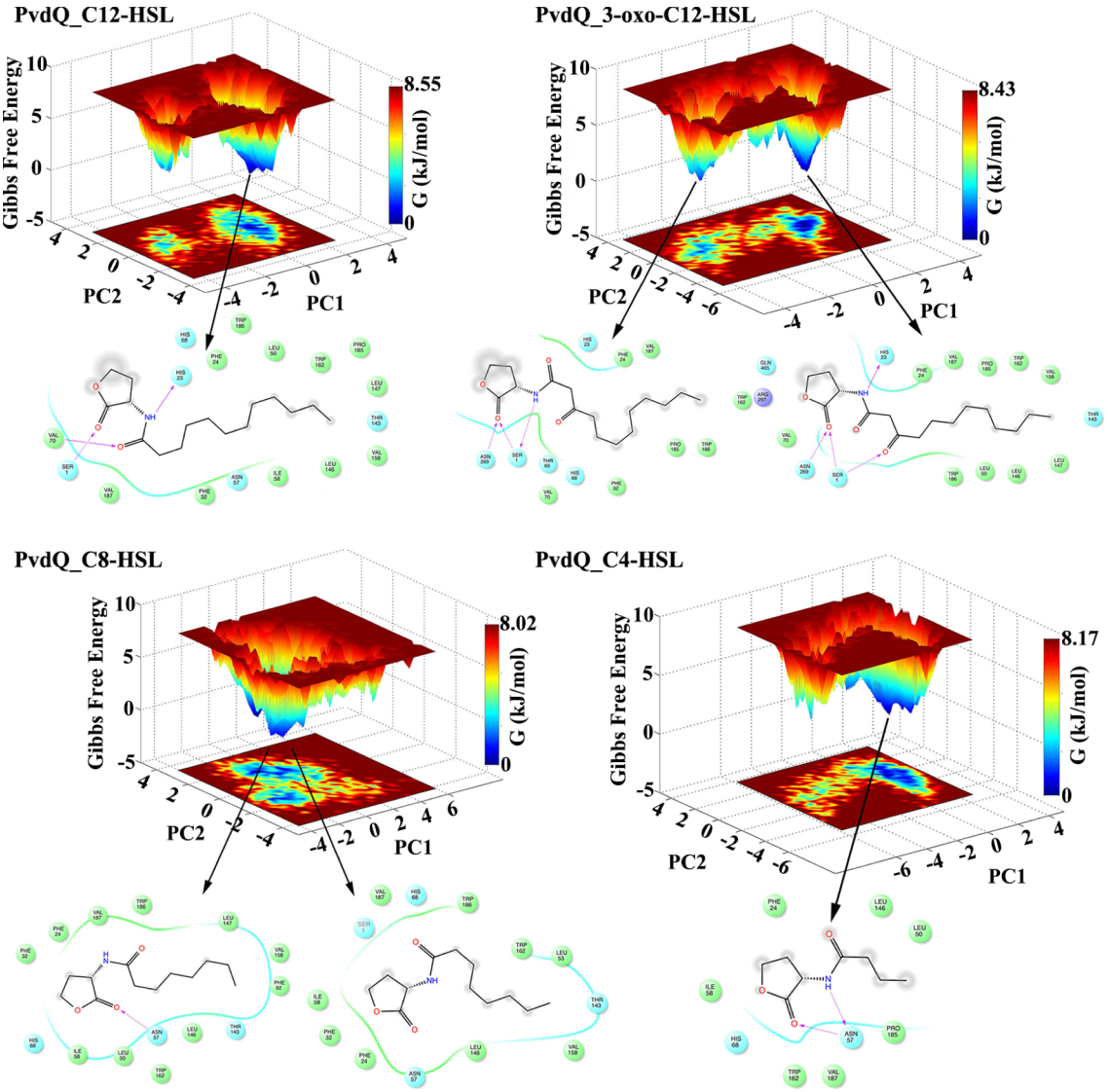

Figure 4.8 The free energy landscapes were generated by projecting the first two principle components of the substrate-bound complexes along with the corresponding minimum energy structures.

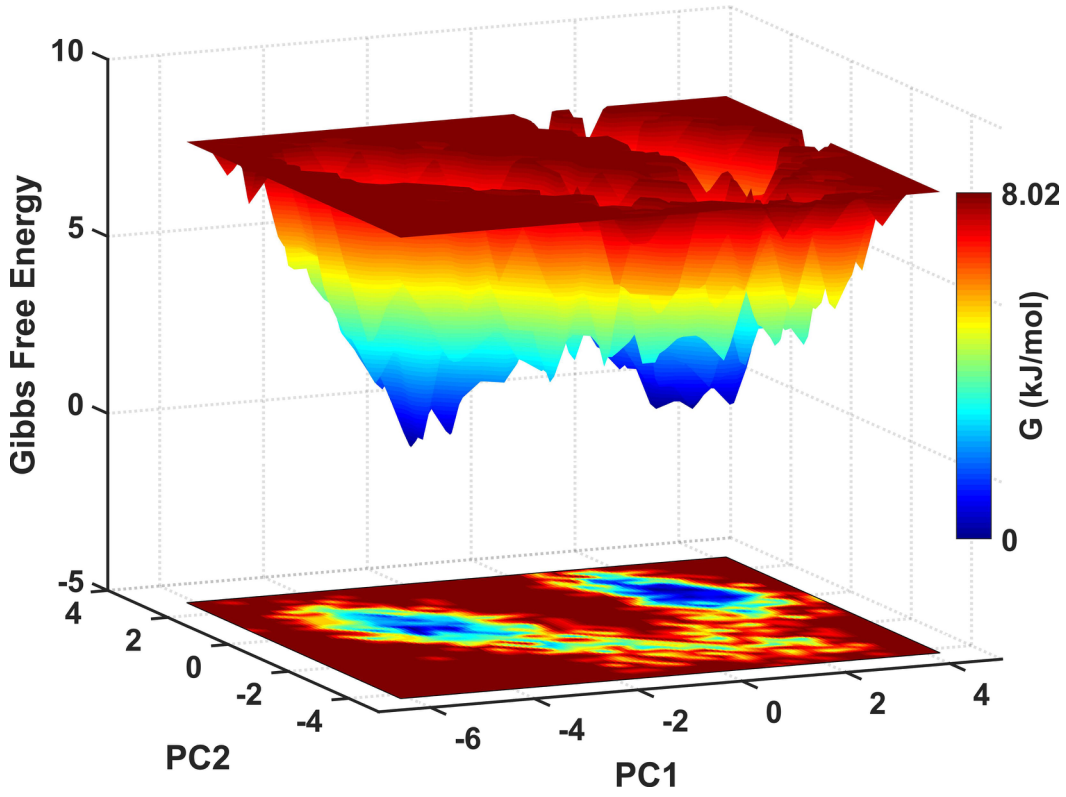

Figure 4.9 The free energy landscape was generated by projecting the first two principle components of the PvdQ protein.

4.3.6 Protein dynamic fingerprint

In order to investigate the dynamical motions of PvdQ induced upon different ligand binding, the protein dynamic fingerprints were calculated. The first principal component was used to analyze the protein dynamic fingerprint since it collectively accounts for around 30% of all dynamical motions. The direction of the arrow indicates the direction of the motion, while the length of the arrow represents the strength of the movement (**Figure 4.10**).

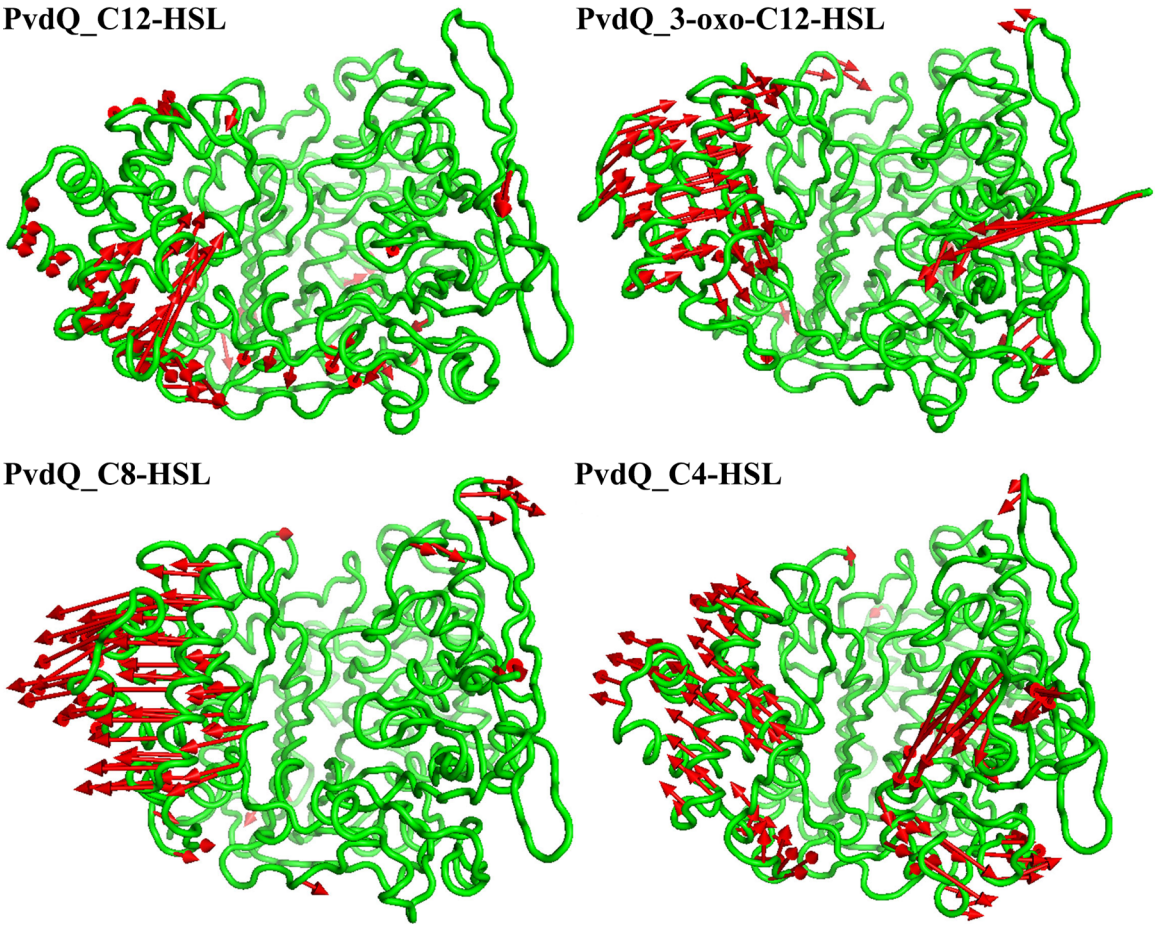

Figure 4.10 Protein dynamic fingerprints for different PvdQ-ligand complexes. The direction of the arrow indicates the direction of the motion, while the length of the arrow representes the strength of the movement.

On the basis of the obtained protein dynamic fingerprint shown, the binding with the C12-HSL ligand reduces the flexibility of the A and B knobs as compared to the unliganded PvdQ and PvdQ bound to other ligands. Therefore, arrangement of the binding pocket of PvdQ remains stable. This is also consistent with the RMSF results. In the 3-oxo-C12-HSL complex, the knobs A and B both move in an opposite direction and closer to each other. These motions push the "gate" Pheβ24 toward its closed state, which

blocks the entrance of the additional 3-oxo substituent from occupying the binding pocket. This supports that this ligand with the extra 3-oxo substituent is not very favored. However, in the C8 and C4-HSL complexes, knobs A and B both move in an opposite direction and further from each other with larger fluctuations. These motions push the "gate" Pheβ24 toward its fully open state, which results in the two ligands being completely buried inside the binding pocket. This explains the disfavoring of the ligand with the short tail group. Therefore, these cooperative movements of the two knobs of PvdQ determine the motions of the binding pocket and are important for correct arrangement of the catalytic site toward substrates.

4.3.7 Residue interaction network and WebPSN analysis

In the interest of further exploring the structural and functional roles of residues, comparative residue interaction network (RIN) analyses were performed using representative structures from different substrate-binding complexes.[244] The residues with high closeness and betweenness centrality are crucial for the stabilization of the protein structure and are likely to play important functional roles.[231,245-247] In **Figure 4.11**, the RIN results illustrate that residues with high closeness and betweenness centrality values are located in the substrate binding pocket. These residues surrounding the substrate are responsible for rapid spreading of information and controlling the interactions of the nodes. However, the differences of the closeness and betweenness centrality values indicate that there exist structural rearrangements in the active sites of different protein-ligand complexes.

In order to further analyze these functionally key residues, the first neighbor residues of

each ligand are retrieved from the interaction network. As shown in **Figure 4.11**, it is clear that the noncovalent interactions in the active site are different for different substrate-binding complexes. C12-HSL substrate has more connection to the surrounding residues and forms a more complicated interaction network (nodes: 44, edges: 99) as compared to other ligands, 3-oxo-C12-HSL (nodes: 41, edges: 88), C8-HSL (nodes: 34, edges: 69), C4-HSL (nodes: 28, edges: 61). The residues Serβ1, Hisβ23, Pheβ24, Metβ30, Pheβ32, Leuβ53, Asnβ57, Thrβ69, Valβ70, Asnβ188, and Asnβ269 in the active site of the C12-HSL complex are the nodes with the highest number of interactions as compared to the other ligands. The van der Waals interaction between Serβ1 and Valβ70, Hisβ23 and Glyβ478, Pheβ24 and Serβ480, Metβ30 and Argβ145, Pheβ32 and Asnβ57, Leuβ53 and Thrα143, Leuβ53 and Leuα146, Asnβ57 and Hisβ68, Valβ70 and Asnβ269, Asnβ188 and Aspβ71 in C12-HSL complex disappear in the other complexes. In the C12-HSL complex, Hisβ23 and Asnβ2, Thrβ69 and Asnβ188 form extra hydrogen bond interactions. These residues are helpful in information communication and regulating the structural rearrangements of the binding pocket.

The decreased number of connections in the 3-oxo-C12-HSL, C8-HSL, and C4-HSL complexes likely indicate that the information spread less and the internal communications of these complexes are decreased accordingly. In particular, the betweenness centrality values of Asnβ269 and Hisβ23 vary among different complexes, which is likely to be important in information communication and biological function of the protein. These analyses of the network communications provide additional information about the roles of key active residues and indicate that the structural rearrangements of the binding pocket affect the binding specificity of the protein.

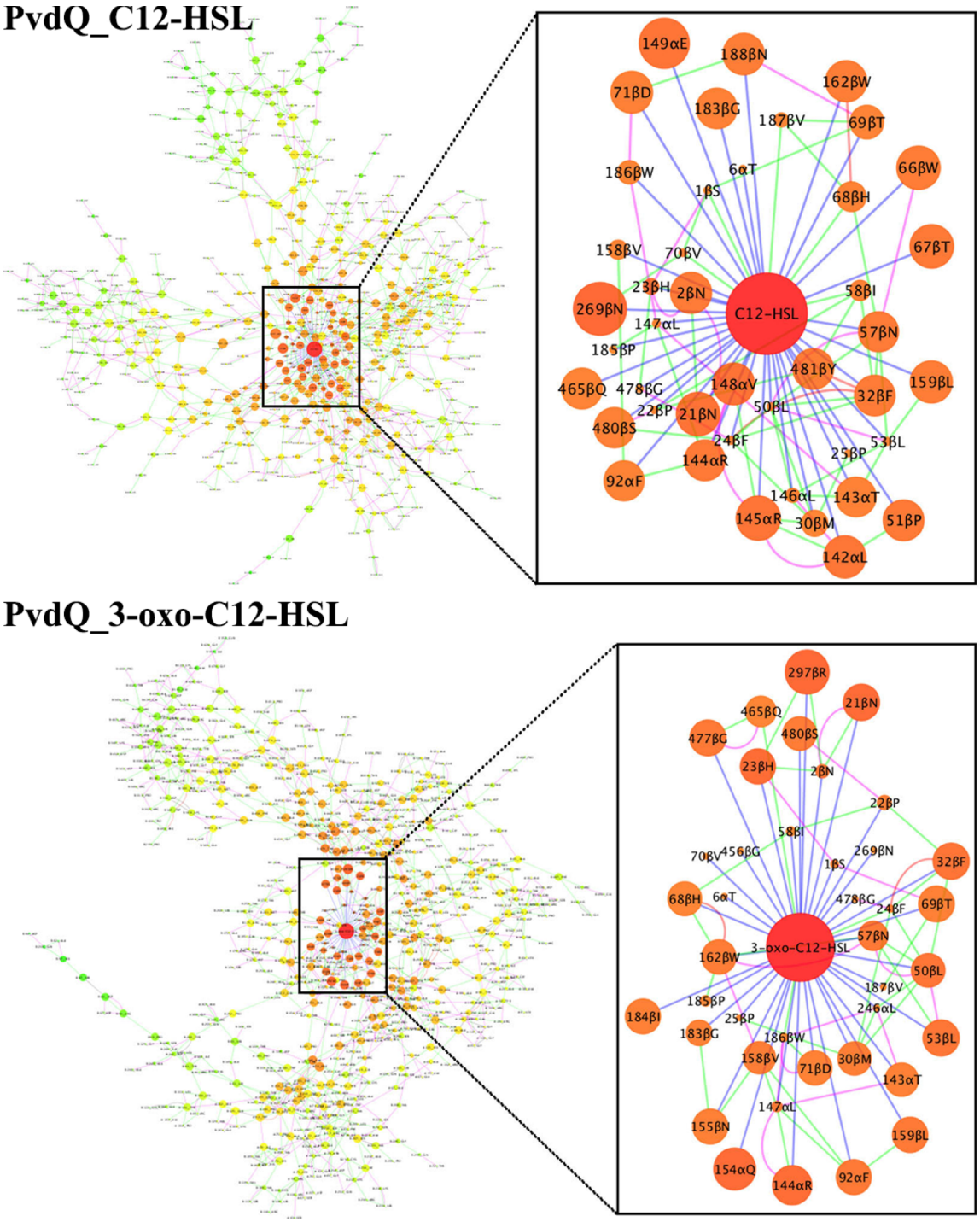

Figure 4.11 RIN of four different ligands binding to the PvdQ. The first neighbor residues of each ligand (right figure) are retrieved from the entire interaction network (left figure). Magenta, green, red, and blue lines represent hydrogen bond, van der Waals, pi-pi stacking interactions, and interaction between closest atoms, respectively.

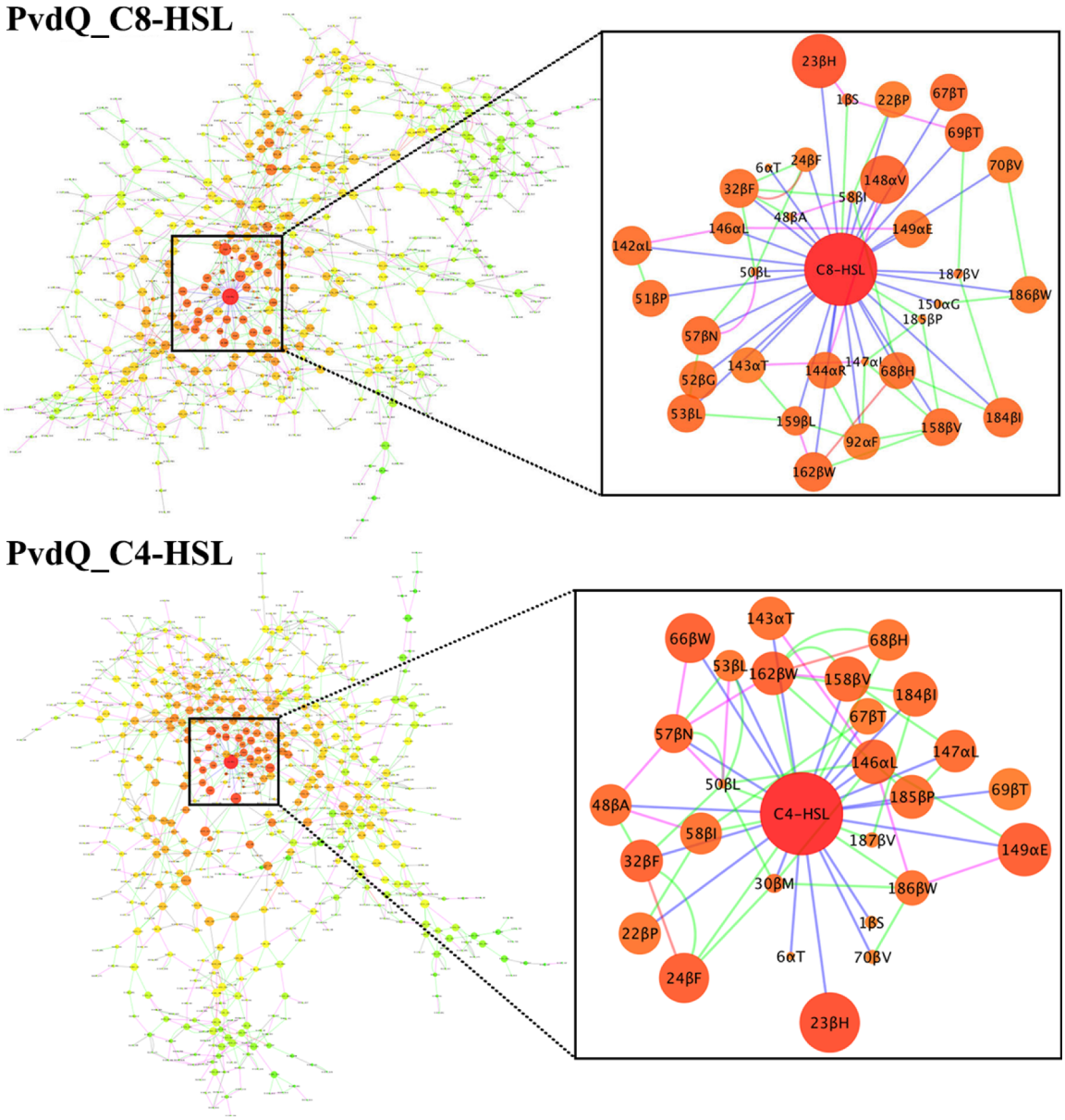

Figure 4.11 (Continued) RIN of four different ligands binding to the PvdQ. The first neighbor residues of each ligand (right figure) are retrieved from the entire interaction network (left figure). Magenta, green, red, and blue lines represent hydrogen bond, van der Waals, pi-pi stacking interactions, and interaction between closest atoms, respectively.

4.3.8 WebPSN analysis

In order to study the global structural communication pathways of each protein-ligand complex, the distinct long-range communication pathways were computed using the

WebPSN server. In **Figure 4.12**, the meta-pathways of each complex demonstrate the potential roles of some key residues and the regions that participated in the stability of the protein.

The C12-HSL complex system has a more complicated global communication pathway, which spreads out over the two knob regions of the protein structure and connects the two chains together. The C12-HSL ligand binding induces changes in the paths resulting in a gain of nodes and links as compared to others, which shows greater signal propagation and potentially results in the ability to minimize the motions of the two knobs. The nodes participating in the most populated communication pathway are shown in **Figure 4.12**. The following nodes are members of the most populated communication pathway: Pheβ370-Trpβ436-Glnβ440-Ileβ451-Valβ475-Trpβ26-Pheβ500-Leuβ498-Glnβ502-Glnβ34-Tyrβ481-Alaβ48-Pheβ32-Pheβ24-Metβ30-Argα145-Gluα49-Thrα52-Argα57. Residues Metβ30, Tyrβ33, Pheβ24 and Pheβ32 are identified as the hubs, which connect the A and B chains together and make links between nodes on the two knobs and the active site. The recurrent residues Trpβ436, Glnβ440, Ileβ451, Valβ475, Trpβ26, Pheβ500 and Glnβ502 are participating in many considered pathways and helpful for information transfer through the protein.

Communication pathways of the 3-oxo-C12-HSL, C8-HSL, and C4-HSL complexes mainly concentrate in the binding site and lose the pathways to the two knobs, which may allow more perturbations in the two knobs upon binding these ligands. For the 3-oxo-C12-HSL complex, it shares some of the same residues in the most populated communication pathway with the C12-HSL complex. However, the communication pathway of the 3-oxo-C12-HSL complex is shorter than that for the C12-HSL complex

and mainly involves residues of the B chain. Residues Glnβ502, Glnβ34, Tyrβ481, Alaβ48, Pheβ24, Metβ30, Argα145, Gluα49, Thrα52, and Argα57 that are peculiar to the C12-HSL complex in signal propagation disappear in the 3-oxo-C12-HSL complex. The following residues are involved in a large number of the communication paths in the 3-oxo-C12-HSL complex and appear to play a large role in structural communication: Trpβ26, Pheβ32, Leuβ50, Asnβ57, Proβ185, Trpβl62, and Pheβ500. For the most populated communication pathway of the C8-HSL complex, it shares the largest number of residues with the 3-oxo-C12-HSL complex. Some recurrent nodes, such as Thrβ479, Hisβ23, Trpβ436, Glnβ440, Ileβ464, Argβ297, Asnβ269, Valβ70, Trpβ186, Aspβ150, Asnβ152, Leuβ153, and Argβ144 are important in the C8-HSL complex communication system. In both the 3-oxo-C12-HSL and C8-HSL complex systems, the communication pathways mainly focus on the residues in the B chain, while residues of the A chain are almost not involved in the pathways.

The communication pathway of the C4-HSL complex shares fewer residues with the other complexes. It forms a short and different communication pathway as compared to others. Residues Ileβ451, Valβ475, Trpβ26, Pheβ500, Tyrβ33, Proα20, Trpα14, Alaα39, Argα40, and Pheα114 play a major part in these communications. The redistributions of these nodes cause the populated communication pathways to involve the two chains together and form a completely different binding mode for the C4-HSL ligand. The communication pathways of each complex demonstrate the structural perturbation caused by different ligand binding and potential roles of some apparently key residues.

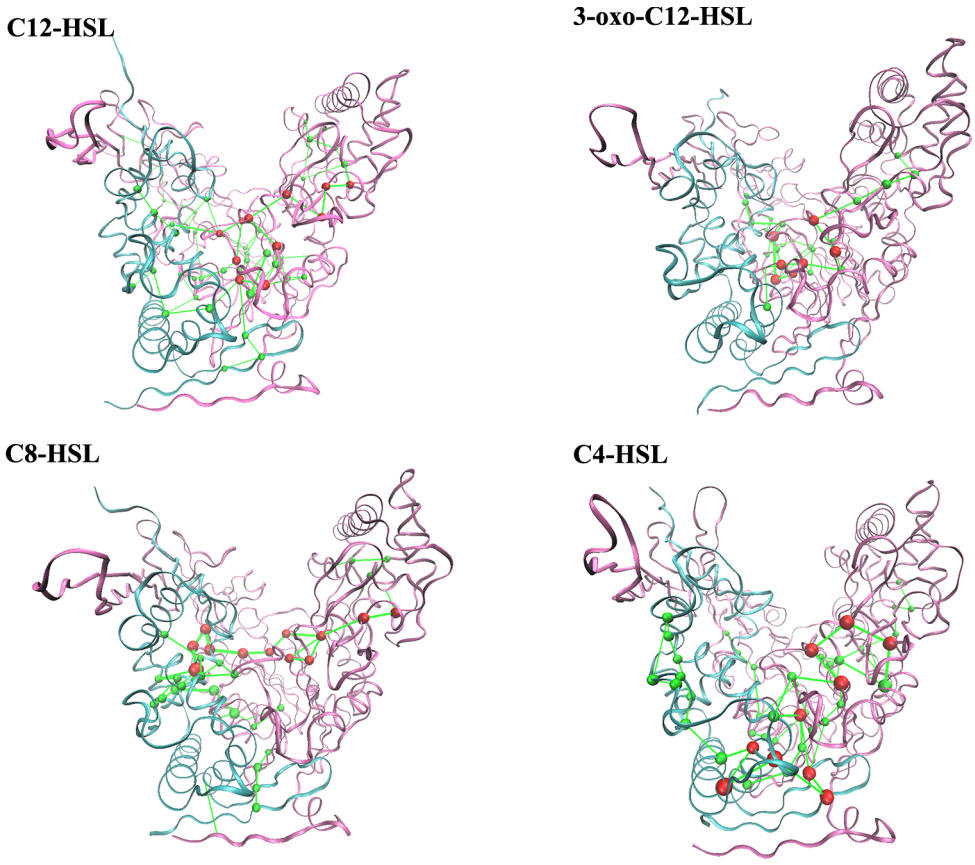

Figure 4.12 Global meta paths of each PvdQ complex achieved by WebPSN and the most populated communication pathways for different complexes. Each spherical node is centered on the Cα-atom of one residue, whose diameter is proportional to the number of links generated by the node. The link thickness is also proportional to link frequencies in the communication paths. The red spherical nodes represent the most populated nodes.

4.3.9 Binding free energy analysis

As a complement to the binding interactions and dynamic motions discussed above, MM-PBSA was used to calculate the binding free energy in order to more deeply understand the energetic changes of PvdQ induced upon different ligand binding (**Table 4.4**). Van der Waals (ΔE_{vdw}) and electrostatic energies (ΔE_{ele}) are favorable for complex formation,

while polar solvation energy (ΔG_{pol}) and entropy ($-T\Delta S$) impair binding free energy. The SASA energy (ΔG_{nonpol}) makes a slight favorable contribution to the total binding free energy. The binding of C12-HSL has the most favorable binding free energy with a value of -12.74 ± 2.17 kcal/mol, which is consistent with the experimental inhibitor binding free energy.[208] The binding free energy of 3-oxo-C12-HSL is unfavorable as compared to that for C12-HSL due to the extra 3-oxo group which changes the van der Waals and polar solvation energy contributions. The change in the electrostatic energy of 3-oxo-C12-HSL may be due to the additional hydrogen bonds between PvdQ and the ligand: Serβ1(HG) – HSL(O4), Hisβ23(H) – HSL(O4), and Serβ1(H1) – HSL(O4). C8-HSL and C4-HSL have the least favorable binding free energies, which are consistent with their low binding affinity with the PvdQ protein.

From the analysis of the individual binding energy components, it is obvious that the van der Waals energy plays a primary driving role in the total binding free energy. The previous inhibitor affinity increased by -1.1 kcal/mol/CH_2, while the predicted binding free energy of the bacterial ligand increases by -0.9 kcal/mol/CH_2. The difference in the binding free energy is due to the enhanced affinity of the inhibitor, which is more potent than structurally similar fatty acids.[208] These results demonstrate that the computationally predicted binding free energy of each PvdQ complex is in agreement with the previous experimental results.[208] A good correlation of predicted binding free energy with *N-acyl* chain length is obtained with a correlation coefficient value of 0.897 (**Figure 4.13**).

Table 4.4 Distribution of the various interaction energies of PvdQ-ligand complexes

MM-PBSA (kcal/mol)				
	C12-HSL	**3-oxo-C12-HSL**	**C8-HSL**	**C4-HSL**
ΔE_{vdw}	-42.76 ± 2.53	-39.96 ± 3.25	-36.80 ± 1.82	-25.17 ± 2.53
ΔE_{ele}	-13.47 ± 2.91	-15.18 ± 4.12	-7.52 ± 1.89	-8.68 ± 2.44
ΔG_{pol}	25.73 ± 2.55	29.07 ± 3.66	22.96 ± 1.98	18.29 ± 2.39
ΔG_{nonpol}	-4.09 ± 0.20	-4.16 ± 0.23	-3.65 ± 0.18	-2.67 ± 0.19
$-T\Delta S$	21.85±1.2	19.39±2.1	15.38±1.5	12.89±2.0
ΔG_{bind}	-12.74 ±2.17	-10.84 ±2.50	-9.63 ±2.11	-5.34 ±2.33

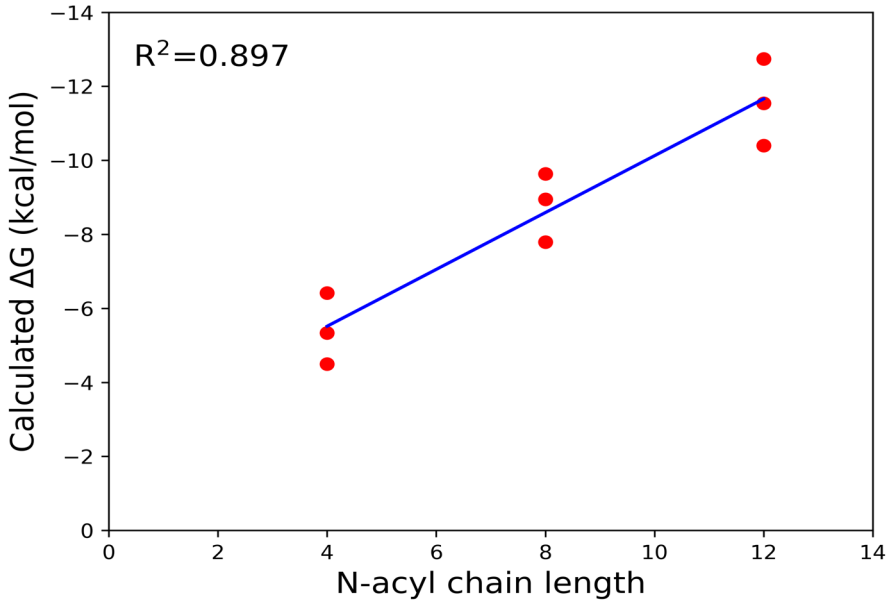

Figure 4.13 Calculated binding free energy with n-acyl chain length.

In order to determine the contribution of each individual residue, the per-residue decomposition energy analysis of the full binding free energy was calculated (**Figure 4.14**). The per-residue decomposition energy analysis contains all parts of the binding

free energy except for the entropy. Residues with a positive interaction energy value impair the binding and vice versa. In this study, in order to focus on the most critical interactions, residues are regarded as playing a key role in substrate binding if their binding free energy is more favorable than -2 kcal/mol.[242-248] There are more key residues favoring the ligand binding in the C12-HSL complex than others. The per-residue decomposition energy results show that Leuα146, Leuα147, Hisβ23, Pheβ24, Leuβ50, Valβ70, Trpβ162, Trpβ186, and Valβ187 make favorable contributions to the binding of the ligands. These key residues are crucial to ligand binding. Residues Trpβ186 and Pheβ24 are the top two contributors for binding in the C12-HSL and 3-oxo-C12-HSL complexes. Trpβ186 is a key residue lining the binding pocket and the movement of the Trpβ186 residue controlls the size the binding pocket. Pheβ24 acts as a gate to separate the solvent and hydrophobic binding pocket. The closed state of Pheβ24 blocks the entrance of 3-oxo-C12-HSL ligand in the binding pocket and the fully open state of Pheβ24 allows the C8-HSL and C4-HSL ligands to be completely buried inside the binding pocket. Valβ70 is the oxyanion hole residue that stabilizes the intermediate catalytic state, Hisβ23 forms polar interactions with the ligand, and Leuα146 is a key residue that moves upon ligand binding. The catalytic center Serβ1 plus residues Asnβ57 and Asnβ269 impair the ligand binding to different extents, which is consistent with the low catalytic efficiency of the PvdQ enzyme. In **Figure 4.15**, the extra hydrophilic 3-oxo substituent of 3-oxo-C12-HSL interacts with surrounding hydrophilic residues, Serβ1 and Hisβ23, and is also repelled by the hydrophobic residues lining the tail binding pocket. Therefore, the 3-oxo-C12-HSL ligand cannot deeply bury itself into the hydrophobic substrate binding pocket. The short acyl-chains of the C8-HSL and C4-HSL are

responsible for the flexibility of these two ligands; therefore, they form different interactional patterns with the surrounding residues. The different residue contributions to the binding of C8-HSL and C4-HSL also demonstrate that C8-HSL and C4-HSL complexes have different binding positions. The binding free energy and per-residue decomposition energy analysis are in agreement with binding specificity and previous experimental results.

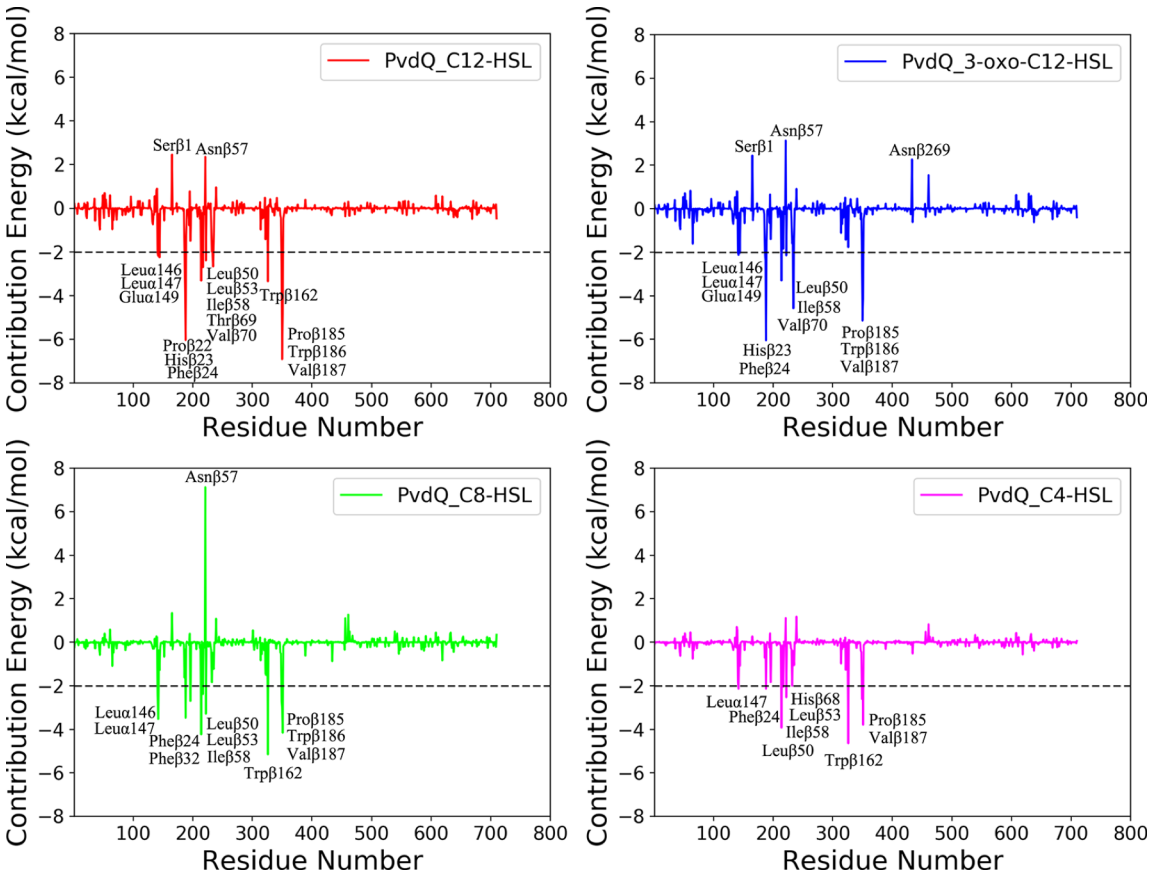

Figure 4.14 Per-residue contribution to the binding free energy of PvdQ in complex with different ligands. Residues with a positive interaction energy value impair the binding and vice versa. Residues are regarded as playing a key role in substrate binding if their binding free energy is more favorable than -2 kcal/mol, which is represented by the dashed line.

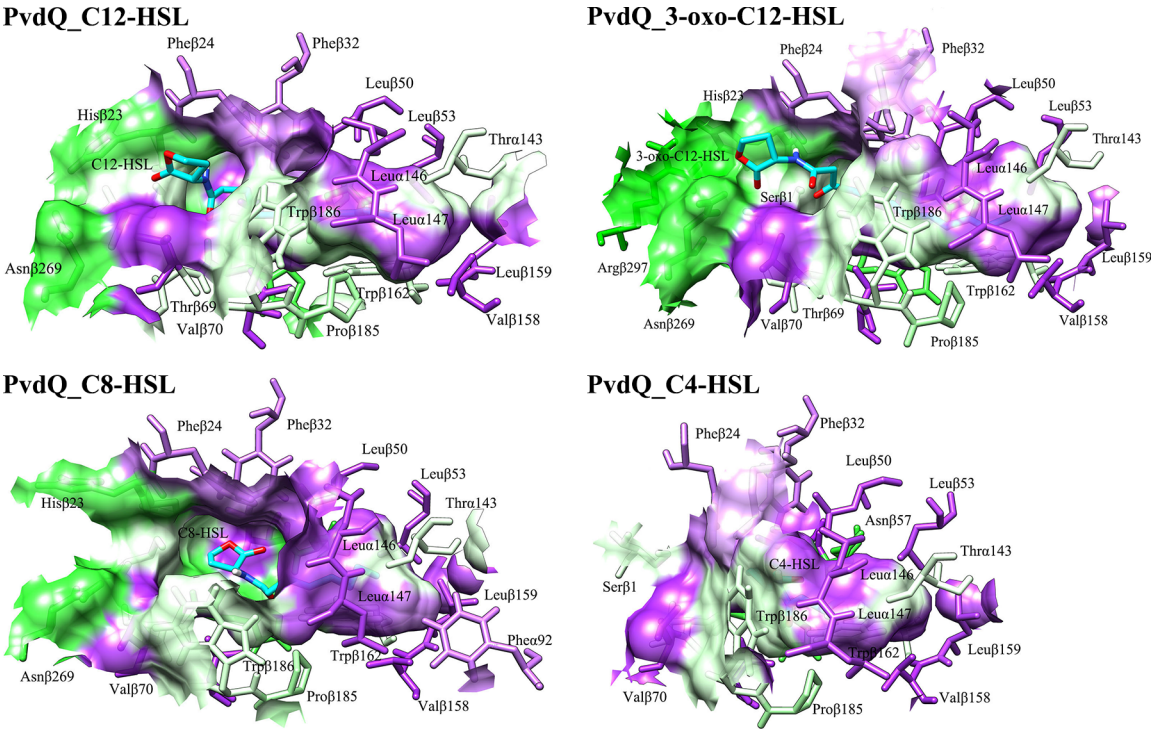

Figure 4.15 The hydrophobicity surface of the substrate binding pocket. Green is the most hydrophilic, while purple is the most hydrophobic (from green to white to purple).

4.4 Conclusions

PvdQ plays an important role in hydrolyzing the AHL ligands, inhibiting quorum sensing and thereby reducing the virulence gene expression and biofilm formation. It is essential to study the structure-dynamic-function relationship of PvdQ with different ligands if we want to improve the catalytic efficiency of PvdQ or design new PvdQ inhibitors. The results of MD simulations from three different docking poses indicated consistent interactions between the protein and ligands. Each of the following residues, Serβ1, Hisβ23, Pheβ24, Metβ30, Pheβ32, Leuβ50, Asnβ57, Thrβ69, Valβ70, Trpβ162, Trpβ186, Asnβ269, Argβ297, and Leuα146 played different key roles in substrate binding

specificity. The global dynamic fingerprint differences of each complex demonstrated that the cooperative movements of the two knobs of PvdQ regulated the motions of the binding pocket and were crucial for correct arrangement of the catalytic site toward substrates. The RIN analysis of each system demonstrated that the C12-HSL substrate formed more connections with surrounding residues as compared to other substrates and the same key residues played a major role in information communication and regulating the structural rearrangements of the binding pocket. The residue communication pathways were predicted for different protein-ligand complexes, which indicated a clear change in the structural communication upon binding of different ligands. The binding of a favorable ligand facilitated the structural communication between the two knobs and the active site. While the binding of the other ligands tended to impair communication between the two knobs and the active site and led to a catalytically inefficient binding site.

The binding of C12-HSL stabilized the PvdQ protein in a minimum energy conformation and produced the lowest binding free energy. The binding affinity was positively correlated with the acyl chain length. The binding free energy calculations of different AHLs were consistent with previous experimental data of the inhibitors. There were polar interactions from the 3-oxo substituent with surrounding residues Serβ1 and Hisβ23 in the lactone head binding pocket, but these interactions of the extra 3-oxo substituent and the closed state of the "gate" Pheβ24 blocked the 3-oxo-C12-HSL from entering into the tail binding pocket. The least favorable ligand was C4-HSL due to its small size and easy diffusion while buried inside the tail binding pocket and was analogous to the C8-HSL binding result from Koch's study.[205] Our data suggested that this buried configuration of

C4-HSL was accessible due to the stabilization by residue Asnβ57. These analyses explained why the PvdQ enzyme does not prefer to hydrolyze the two major endogenously generated communication molecules of *P. aeruginosa*, C4-AHL or 3-oxo-C12-AHL. This study provided a molecular basis for better understanding the structure-dynamic-function relationship of the PvdQ enzyme and demonstrated determinants of substrate recognition and discrimination.

Chapter 5 Future directions

The PvdQ enzyme has different preferences for *N*-acyl substrates with different acyl chain lengths and substituents. The results from this study provides a molecular foundation for better understanding the structure-dynamic-function relationship of the PvdQ protein and demonstrates the determinants of different ligand recognition and discrimination. Our results indicated that each of the following residues, Serβ1, Hisβ23, Pheβ24, Metβ30, Pheβ32, Leuβ50, Asnβ57, Thrβ69, Valβ70, Trpβ162, Trpβ186, Asnβ269, Argβ297, and Leuα146 play different key roles in substrate binding specificity. Our work also showed that the closed state of Pheβ24 blocks the entrance of the 3-oxo-C12-HSL ligand in the binding pocket while the fully open state of Pheβ24 allows the C8-HSL and C4-HSL ligands to be completely buried inside the binding pocket. Therefore, further analysis of the functional effects of mutations at specific positions will aid in obtaining more information to improve the catalytic efficiency of PvdQ. Our study also provided a platform to develop inhibitors to limit the function of PvdQ, and design a more efficient and promiscuous PvdQ to inhibit the growth of specific or a wide range of bacteria.

Insight into the residue Serβ1 mutations to improve the catalytic efficiency of the PvdQ.

As discussed in this study, residue Serβ1 of the PvdQ protein is the nucleophilic catalytic residue, which can deprotonate its own hydroxyl group to activate its catalytic role.

However, PvdQ does not have strong catalytic efficiency. Therefore, it would be innovative to mutate the catalytic center Serβ1 and/or residues in the vicinity to improve the catalytic efficiency. We hypothesize that the mutation of the Serβ1 to Cysβ1 would improve the catalytic efficiency of the N-terminal nucleophile. The side chain of Serβ1 must be deprotonated by its own N-terminal amine. Then the activated Serβ1 nucleophile attacks the carbonyl carbon of the scissile bond of the substrate and results in the formation of the first tetrahedral transition state and acyl-enzyme intermediate. Cysteine, with its more acidic side chain, is more easily deprotonated than serine. Cysteine occurs in various enzymes where it is actively involved in chemical reactions, such as, viral C3 protease and papain super-families. The mutation of the Serβ1 into Cysβ1 would make the N-terminus more acidic before substrate binding and improve the nucleophilic catalysis.

Investigating the impact of residue Pheβ24 mutations on the binding efficiency of PvdQ protein toward different ligands.

This study provided a molecular basis for design of more effecient PvdQ that can hydrolyze the ligands with 3-oxo substituent. Residue Pheβ24 acts as a gate to separate the solvent from the hydrophobic binding pocket. Our results indicated that the closed state of Pheβ24 blocks the entrance of 3-oxo-C12-HSL ligand into the binding pocket. It would be interesting to study the effect of Pheβ24 mutation on the binding efficiency of PvdQ toward different ligands. Residue Pheβ24 can be selected to be substituted by all other 19 possible amino acids though *in silico* mutagenesis. After energy minimization, the new models will be used to dock different ligands to generate energetically favorable

docked poses for further analysis. Large-scale change of the side chain of Pheβ24 could make some changes of the size of the entrance of the binding pocket and binding property of the PvdQ protein.

Study the impact of PvdQ binding sites mutations on the size of the binding pocket to accommodate different ligands.

Our results showed that the entire structures of the C8-HSL and C4-HSL ligands were completely buried inside the binding pocket. The polar head group was surrounded by unfavorable hydrophobic residues and the shorter tail group contributed to the lower hydrophobic energy. The distance between the Oγ atom of the catalytic center Serβ1 and the carbonyl carbon atom of the amide bond of C8-HSL and C4-HSL are far beyond hydrogen bonding distance. It is meaningful to study the mutations of the binding sites to change the size of the binding pocket to accommodate different sizes of ligands. The modification of the shape and electrostatics of each binding site residues to adapt the ligand binding would shed some light on improving the catalytic efficiency of PvdQ protein toward different ligands.

Develop inhibitors to block PvdQ action in iron acquisition

Since PvdQ plays a key role in siderophore biosynthesis in order to maintain iron homeostasis, it could be beneficial to develop inhibitors to block PvdQ action in order to disrupt iron acquisition, thereby impairing bacterial viability. The PvdQ enzyme tolerates different head groups but is more selective for the length and functionalization of the acyl chain group. Therefore, a structure-based drug design method can be used to discover and optimize new inhibitors. This work demonstrated that the PvdQ enzyme prefers ligands

with long *N*-acyl chain group. Modifying the head group of the PvdQ ligand while keeping the length of the acyl tail group will provide some new clues toward structure-based drug design. Therefore, the design new inhibitors of PvdQ that can inhibit the iron acquisition and bacterial growth appears promising and can be accomplished by using the molecular information obtained from this study.

Design of a functional molecule to inhibit bacterial growth.

It is beneficial to design a functional organic catalyst to degrade the quorum signaling molecules. We hypothesize that the organic catalyst mimic that has a similar functional group arrangement as the PvdQ enzyme would hydrolyze the quorum signaling molecules. The functional groups should include a nucleophilic group, a general base to promote the intermediate state, and a general acid to stabilize the protonated general base, thereby protonating the leaving group and leading to an acyl-enzyme intermediate. Further cleavage of the acyl-enzyme intermediate can be mediated by water acting as the attacking nucleophile; then completing the process with regeneration of the artificial enzyme. Therefore, it is compelling to incorporate these functional groups, which have the similar function and orientation with the active sites of the PvdQ protein, on a potential binding scaffold, such as a modified cyclodextrin or crown ether. Then the starting structure can be obtained followed by an energy minimization procedure. The molecular docking and MD simulations can be used to calculate the binding energy and characterize the specific interaction to acquire the stable states in the active site. The enzyme-substrate pose needs to be scored and refined. The top scoring poses should be re-docked and mutations are needed to enhance the catalytic efficiency. The above

procedures need to be employed iteratively to obtain the final optimized molecular structure. The final organic AHL degradation catalyst will be made and tested with bacterial signaling molecules, which can be applied in a water system, hospital settings, and the International Space Station.

Chapter 6 References

1. Kipnis E., Sawa T., and Wiener-Kronish J. Targeting mechanisms of Pseudomonas aeruginosa pathogenesis. Med. Mal. Infect. 2006, 36, 78–91.
2. Obritsch M. D., Fish D. N., MacLaren R., and Jung R. National surveillance of antimicrobial resistance in Pseudomonas aeruginosa isolates obtained from intensive care unit patients from 1993 to 2002. Antimicrob. Agents Chemother. 2004, 48, 4606–4610.
3. Friedland I., Gallagher G., King T., and Woods G. L. Antimicrobial susceptibility patterns in Pseudomonas aeruginosa: Data from a multicenter intensive care unit surveillance study (iss) in the United States. J. Chemother. 2004, 16(5), 437–441.
4. Chastre J., and Fagon J. Y. Ventilator-associated pneumonia. Am. J. Respir. Crit. Care Med. 2002, 165, 867–903.
5. Chastre J., and Trouillet J. L. Problem pathogens (Pseudomonas aeruginosa and acinetobacter). Semin. Respir. Infect. 2000, 15, 287–298.
6. Rello J., Rue M., Jubert P., Muses G., Sonora R., Valles J., and Niederman M. S. Survival in patients with nosocomial pneumonia: Impact of the severity of illness and the etiologic agent. Crit. Care Med. 1997, 25, 1862–1867.
7. Crouch B. S., Wunderink R. G., Jones C. B., and Leeper Jr. K. V. Ventilator-associated pneumonia due to Pseudomonas aeruginosa. Chest. 1996, 109, 1019–1029.
8. Sadikot R. T., Blackwell T. S., Christman J. W., and Prince A. S. Pathogen-host interactions in Pseudomonas aeruginosa pneumonia. Am. J. Respir. Crit. Care

Med. 2005, 171, 1209–1223.

9. Kollef M. H., Chastre J., Fagon J. Y. *et al*. Global prospective epidemiologic and surveillance study of ventilator-associated pneumonia due to Pseudomonas aeruginosa. Crit. Care Med. 2014, 42, 2178–2187.
10. Takesue Y., Watanabe A., Hanaki H., Kusachi, S. *et al*. Nationwide surveillance of antimicrobial susceptibility patterns of pathogens isolated from surgical site infections (SSI) in Japan. J. Infect. Chemother. 2012, 18, 816–826.
11. Ortega M., Marco F., Soriano A., Almela M., Martinez J. A., Pitart C., and Mensa J. Epidemiology and prognostic determinants of bacteraemic catheter-acquired urinary tract infection in a single institution from 1991 to 2010. J. Infect. 2013, 67, 282–287.
12. Emerson J., Rosenfeld M., McNamara S., Ramsey B., and Gibson R. L. Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis. Pediatr. Pulmonol. 2002, 34, 91–100.
13. Son M. S., Matthews W. J., Kang Y., Nguyen D. T., and Hoang T. T. In vivo evidence of Pseudomonas aeruginosa nutrient acquisition and pathogenesis in the lungs of cystic fibrosis patients. Infect. Immun. 2007, 75, 5313–5324.
14. Sorger-Domenigg T. Novel Pseudomonas Aeruginosa Small Regulatory RNAs. PhD thesis, University of Vienna, Vienna. 2010.
15. Kollef M. H. Prevention of hospital-associated pneumonia and ventilator-associated pneumonia. Crit. Care Med. 2004, 32, 1396–1405.
16. Hoffken G., and Niederman M. S. Nosocomial pneumonia: The importance of a de-escalating strategy for antibiotic treatment of pneumonia in the icu. Chest.

2002, 122, 2183–2196.

17. Warren D. K., Hill H. A., Merz L. R., Kollef M. H., Hayden M. K., Fraser V. J. *et al*. Cycling empirical antimicrobial agents to prevent emergence of antimicrobial-resistant gram-negative bacteria among intensive care unit patients. Crit. Care Med. 2004, 32, 2450–2456.
18. Burgess D. S. Use of pharmacokinetics and pharmacodynamics to optimize antimicrobial treatment of Pseudomonas aeruginosa infections. Clin. Infect. Dis. 2005, 40, S99–S104.
19. Iregui M., Ward S., Sherman G., Fraser V. J., and Kollef M. H. Clinical importance of delays in the initiation of appropriate antibiotic treatment for ventilator-associated pneumonia. Chest. 2002, 122, 262–268.
20. Klepser M. E. Role of nebulized antibiotics for the treatment of respiratory infections. Curr. Opin. Infect. Dis. 2004, 17, 109–112.
21. Feldman M., Bryan R., Rajan S., Scheffler L., Brunnert S., Tang H., and Prince A. Role of flagella in pathogenesis of Pseudomonas aeruginosa pulmonary infection. Infect. Immun. 1998, 66, 43–51.
22. Wick M. J., Hamood A. N., and Iglewski B. H. Analysis of the structure-function relationship of Pseudomonas aeruginosa exotoxin A. Mol. Microbiol. 1990, 4, 527–535.
23. Pathak A. New Vaccines for Infectious Diseases: Immunological Targeting of the Quorum Sensing System of Pseudomonas aeruginosa. PhD thesis, UK University of Nottingham, Nottingham. 2012.
24. Gambello M. J., Kaye S., and Iglewski B. H. LasR of Pseudomonas aeruginosa is

a transcriptional activator of the alkaline protease gene (apr) and an enhancer of exotoxin A expression. Infect. Immun. 1993, 61, 1180–1184.

25. Toder D. S., Ferrell S. J., Nezezon J. L., Rust L., and Iglewski B. H. lasA and lasB genes of Pseudomonas aeruginosa: analysis of transcription and gene product activity. Infect. Immun. 1994, 62, 1320–1327.
26. Nakae T. Role of membrane permeability in determining antibiotic resistance in Pseudomonas aeruginosa. Microbiol. Immunol. 1995, 39, 221–229.
27. Nikaido H. Multidrug efflux pumps of gram-negative bacteria. J. Bacteriol. 1996, 178, 5853–5859.
28. Piddock L. J. Multidrug-resistance efflux pumps-not just for resistance. Nat. Rev. Microbiol. 2006, 4, 629–636.
29. Tang Y., Ali Z., Zou J., Jin G., Zhu J., Yanga J. *et al*. Detection methods for Pseudomonas aeruginosa: history and future perspective. RSC Adv. 2017, 7, 51789–51800.
30. Franzetti L., and Scarpellini M. Characterisation of Pseudomonas spp. isolated from foods. Ann. Microbiol. 2007, 57, 39–47.
31. Joyanes P., del Carmen C. M., Martinez-Martinez L., and Perea E. J. Evaluation of the VITEK 2 system for the identification and susceptibility testing of three species of nonfermenting gram-negative rods frequently isolated from clinical samples. J. Clin. Microbiol. 2001, 39, 3247–3253.
32. Funke G., and Funke-Kissling P. Use of the BD PHOENIX Automated Microbiology System for direct identification and susceptibility testing of gram-negative rods from positive blood cultures in a three-phase trial. J. Clin.

Microbiol. 2004, 42, 1466–1470.

33. Rüger M., Ackermann M., and Reichl U. Species-specific viability analysis of Pseudomonas aeruginosa, Burkholderia cepacia and Staphylococcus aureus in mixed culture by flow cytometry. BMC Microbiol. 2014, 14, 14–56.

34. Rao A. R., Splaingard M. S., Gershan W. M., Havens P. L., Thill A., and Barbieri J. T. Detection of Pseudomonas aeruginosa type III antibodies in children with tracheostomies. Pediatr. Pulmonol. 2005, 39, 402–407.

35. Choi H. J. *et al*. Improved PCR for identification of Pseudomonas aeruginosa. Appl. Microbiol. Biotechnol. 2013, 97, 3643–3651.

36. Kingsford N. M., and Raadsma H. W. Detection of Pseudomonas aeruginosa from ovine fleece washings by PCR amplification of 16S ribosomal RNA. Vet. Microbiol. 1995, 47, 71–70.

37. De Vos D., Lim Jr. A., Pirnay J. P., Struelens M., Vandenvelde C., Duinslaeger L., Vanderkelen A., and Cornelis P. Direct detection and identification of Pseudomonas aeruginosa in clinical samples such as skin biopsy specimens and expectorations by multiplex PCR based on two outer membrane lipoprotein genes, oprI and oprL. J. Clin. Microbiol. 1997, 35, 1295–1299.

38. Qin X., Emerson J., Stapp J., Stapp L., Abe P., and Burns J. L. Use of real-time PCR with multiple targets to identify Pseudomonas aeruginosa and other nonfermenting gram-negative bacilli from patients with cystic fibrosis. J. Clin. Microbiol. 2003, 41, 4312–4317.

39. Motoshima M., Yanagihara K., Fukushima K. *et al*. Rapid and accurate detection of Pseudomonas aeruginosa by real-time polymerase chain reaction with melting

curve analysis targeting gyrB gene. Diagn. Microbiol. Infect. Dis. 2007, 58, 53–58.

40. Abou Tayoun A. N., Burchard P. R., Caliendo A. M., Scherer A., and Tsongalis G. J. A multiplex PCR assay for the simultaneous detection of Chlamydia trachomatis, Neisseria gonorrhoeae, and Trichomonas vaginalis. Exp. Mol. Pathol. 2015, 98, 214–218.
41. Kurupati P., Kumarasinghe G., and Laa Poh C. Direct identification of Pseudomonas aeruginosa from blood culture bottles by PCR-enzyme linked immunosorbent assay using oprI gene specific primers. Mol. Cell. Probes. 2005, 19, 417–421.
42. Dhama K., Karthik K., Chakraborty S. *et al.* Loop-mediated isothermal amplification of DNA, (LAMP)-a new diagnostic tool lights the world of diagnosis of animal and human pathogens: a review. Pak. J. Biol. Sci. 2014, 17, 151–166.
43. Parmar H., Dholakia A., Vasavada D., and Singhala H. The Current status of Antibiotic sensitivity of Pseudomonas aeruginosa isolated from various clinical samples. Int. J. Res. Med. 2013, 2, 1–6.
44. Mesaros N., Nordman P., Plesiat P., Roussel-Delvallez M., Van Eldere J., Glupczynski Y., Van Lathem .Y, Jacobs F., Lebecque P., Malfroot A., Tulkens P. M. and Van Bambeke F. Pseudomonas aeruginosa: resistance and therapeutic options at the turn of the new millennium. Clin. Microbiol. infect. 2007, 13, 560–578.
45. Lister P. D., Wolter D. J., and Hanson N. D. Antibacterial-Resistant Pseudomonas

aeruginosa: Clinical Impact and Complex Regulation of Chromosomally Encoded Resistance Mechanisms. Clin. Microbiol. Rev. 2009, 22, 582–610.

46. Abdulla A. M. H. Molecular Characterization of Multidrug Resistant Pseudomonas Aeruginosa. PhD thesis, University of Malaya, Kuala Lumpur, Malaysia. 2012.

47. Lister P. D. β-Lactamase Inhibitor Combinations with Extendedmpur, Mal Penicillins: Factors Influencing Antibacterial Activity against Enterobacteriaceae and Pseudomonas aeruginosa. Pharmacotherapy. 2000, 20, 213S–218S.

48. Lodise Jr. T. P., Lomaestro B., and Drusano G. L. Piperacillin-tazobactam for Pseudomonas aeruginosa infection: clinical implications of an extended-infusion dosing strategy. Clin. Infect. Dis. 2007, 44, 357–363.

49. Zuanazzi D., Souto R., Mattos M., Zuanazzi M., Tura B., Sansone C. *et al*. Prevalence of potential bacterial respiratory pathogens in the oral cavity of hospitalised individuals. Arch. Oral. Biol. 2010, 55, 21–28.

50. Rizek C., Fu L., Dos Santos L. C., Leite G., Ramos J., Rossi F., Guimaraes T., Levi A. S., and Costa S. F. Characterization of carbapenem-resistant Pseudomonas aeruginosa clinical isolates, carrying multiple genes coding for antibiotic resistance. Ann. Clin. Microbiol. and Antimicrob. 2014, 13, 1–5.

51. Carmeli Y., Troillet N., Eliopoulos G. M., and Samore M. H. Emergence of antibiotic-resistant Pseudomonas aeruginosa: comparison of risks associated with different antipseudomonal agents. Antimicrob Agents Chemother. 1999, 43, 1379–1382.

52. Jones R. N., Kirby J. T., Beach M. L., Biedenbach D. J., and Pfaller M. A.

Geographic variations in activity of broad-spectrum β-lactams against Pseudomonas aeruginosa: summary of the worldwide SENTRY Antimicrobial Surveillance Program (1997–2000). Diagn. Microbiol. Infect. Dis. 2002, 43, 239–243.

53. Ennis D. M., and Cobbs C. G. The newer cephalosporins. Aztreonam and imipenem. Infect. Dis. Clin. North Am. 1995, 9, 687–713.

54. Giamarellou H., and Kanellakopoulou K. Current therapies for Pseudomonas aeruginosa. Crit. Care Clin. 2008, 24, 261–278.

55. Falagas M. E., Kasiakou S. K., and Saravolatz L. D. Colistin: the revival of polymyxins for the management of multidrug-resistant gram-negative bacterial infections. Clin. Infect. Dis. 2005, 40, 1333–1341.

56. Michalopoulos A., and Falagas, M. E. Colistin and polymyxin B in critical care. Crit. Care Clin. 2008, 24, 377–391.

57. Kasiakou S. K., Michalopoulos A., Soteriades E. S., Samonis G., Sermaides G. J., and Falagas, M. E. Combination therapy with intravenous colistin for management of infections due to multidrug-resistant Gram-negative bacteria in patients without cystic fibrosis. Antimicrob Agents Chemother. 2005, 49, 3136–3146.

58. Lang B. J., Aaron S. D., Ferris W., Hebert P. C., and Macdonald N. E. Multiple Combination Bactericidal Antibiotic Testing for Patients with Cystic Fibrosis Infected with Multiresistant Strains of Pseudomonas aeruginosa. Am. J. Respir. Crit. Care Med. 2000, 162, 2241–2245.

59. Ramsey, B. W. Management of pulmonary disease in patients with cystic fibrosis.

N. Engl. J. Med. 1996, 335, 179–188.

60. Ratjen F., Comes G., Paul K., Posselt H. G., Wagner T. O. F., and Harms K. Effect of continuous antistaphylococcal therapy on the rate of P. aeruginosa acquisition in patients with cystic fibrosis. Pediatr. Pulmonol. 2001, 31, 13–16.

61. Gibson R. L., Emerson J., McNamara S. *et al*. Significant Microbiological Effect of Inhaled Tobramycin in Young Children with Cystic Fibrosis. Am. J. Respir. Crit. Care Med. 2003, 167, 841–849.

62. Barclay M. L., Begg E. J., Chambers S. T., Thornley P. E., Pattemore P. K., and Grimwood K. Adaptive resistance to tobramycin in Pseudomonas aeruginosa lung infection in cystic fibrosis. J. Antimicrob. Chemother. 1996, 37, 1155–1164.

63. Conway S. P., Pond M. N., Watson A., Etherington C., Robey H. L., and Goldman M. H. Intravenous colistin sulphomethate in acute respiratory exacerbations in adult patients with cystic fibrosis. Thorax. 1997, 52, 987–993.

64. Micek, S. T., Lloyd, A. E., Ritchie, D. J., Reichley, R. M., Fraser, V. J., and Kollef, M. H. Pseudomonas aeruginosa bloodstream infection: importance of appropriate initial antimicrobial treatment. Antimicrob Agents Chemother. 2005, 49, 1306–1311.

65. Yabuuchi E., Kosako Y., Oyaizu H., Yano I., Hotta H., Hashimoto Y., Ezaki T., and Arakawa M. Proposal of Burkholderia gen. nov. and transfer of seven species of the genus Pseudomonas homology group II to the new genus, with the type species Burkholderia cepacia (Palleroni and Holmes 1981) comb. nov. Microbiol. Immunol. 1992, 36, 1251–1275.

66. Vandamme P., Holmes B., Coenye T., Goris J., Mahenthiralingam E., LiPuma J.

J., and Govan J. R. Burkholderia cenocepacia sp. nov.-a new twist to an old story. Res. Microbiol. 2003, 154, 91–96.

67. Vandamme P., Holmes B., Vancanneyt M., Coenye T., Hoste B., Coopman R., Revets H., Lauwers S., Gillis M., Kersters K., and Govan J. R. Occurrence of multiple genomovars of Burkholderia cepacia in cystic fibrosis patients and proposal of Burkholderia multivorans sp. nov. Int. J. Syst. Evol. Microbiol. 1997, 47, 1188–1200.

68. Vanlaere E., Baldwin A., Gevers D., Henry D., De Brandt E., Lipuma J. J., Mahenthiralingam E., Speert D. P., Dowson C., and Vandamme P. Taxon K, a complex within the Burkholderia cepacia complex, comprises at least two novel species, Burkholderia contaminans sp. nov. and Burkholderia lata sp. nov. Int. J. Syst. Evol. Microbiol. 2009, 59, 102–111.

69. Vanlaere E., Lipuma J. J., Baldwin A., Henry D., De Brandt E., Mahenthiralingam E., Speert D., Dowson C., and Vandamme P. Burkholderia latens sp. nov., Burkholderia diffusa sp. nov., Burkholderia arboris sp. nov., Burkholderia seminalis sp. nov. and Burkholderia metallica sp. nov., novel species within the Burkholderia cepacia complex. Int. J. Syst. Evol. Microbiol. 2008, 58, 1580–1590.

70. Isles A., Maclusky I., Corey M., Gold R., Prober C., Fleming P., and Levison, H. Pseudomonas cepacia infection in cystic fibrosis: an emerging problem. J. Pediatr. 1984, 104, 206–210.

71. Urban T. A., Griffith A., Torok A. M., Smolkin M. E., Burns J. L., and Goldberg J. B. Contribution of Burkholderia cenocepacia flagella to infectivity and

inflammation. Infect. Immun. 2004, 72, 5126–34.

72. Visser M. B., Majumdar S., Hani E., and Sokol P. A. Importance of the omibactin and pyochelin siderophore transport systems in Burkholderia cenocepacia lung infections. Infect. Immun. 2004, 72, 2850–2857.

73. Conway B. A., Venu V., and Speert D. P. Biofilm formation and acyl homoserine lactone production in the Burkholderia cepacia complex. J. B. acteriol. 2002, 184, 5678–5685.

74. Venturi V., Friscina A., Bertani I., Devescovi G., and Aguilar C. Quorum sensing in the Burkholderia cepacia comple. Res. Microbiol. 2004, 155, 238–244.

75. Glendinning K. J., Parsons Y. N., Duangsonk K., Hales B. A., Humphreys D., Hart C. A., and Winstanley C. Sequence divergence in type III secretion gene clusters of the Burkholderia cepacia complex. FEMS Microbiol. Lett. 2004, 235, 229–235.

76. Corbett C. R., Burtnick M. N., Kooi C., Woods D. E., and Sokol P. A. An extracellular zinc metalloprotease gene of Burkholderia cepacia. Microbiol. 2003, 149, 2263–2271.

77. Porter L. A., and Goldberg J. B. Influence of neutrophil defects on *Burkholderia cepacia* complex pathogenesis. Front. Cell. Infect. Microbiol. 2011, 1, 9.

78. Tomich M., Herfst C.A., Golden J.W., and Mohr C.D. Role of flagella in host cell invasion by Burkholderia cepacia. Infect. Immun. 2002, 70, 1799–1806.

79. Kooi C., Subsin B., Chen R., Pohorelic B., and Sokol P. A. Burkholderia cenocepacia ZmpB is a broad-specificity zinc metalloprotease involved in virulence. Infect. Immun. 2006, 74, 4083–4093.

80. Bamford S., Ryley H., and Jackson S. K. Highly purified lipopolysaccharides from Burkholderia cepacia complex clinical isolates induce inflammatory cytokine responses via TLR4-mediated MAPK signalling pathways and activation of NFKB. Cell. Microbiol. 2007, 9, 532–543.

81. Ortega X., Hunt T. A., Loutet S., Vinion-Dubiel A. D., Datta A., Choudhury B., Goldberg J. B., Carlson R., and Valvano M. A. Reconstitution of O-specific lipopolysaccharide expression in Burkholderia cenocepacia strain J2315, which is associated with transmissible infections in patients with cystic fibrosis. J. Bacteriol. 2005, 187, 1324–1333.

82. Saldías M. S., Ortega X., and Valvano M. A. Burkholderia cenocepacia O antigen lipopolysaccharide prevents phagocytosis by macrophages and adhesion to epithelial cells. J. Med. Microbiol. 2009, 58, 1542–1548.

83. Thomas M. Iron acquisition mechanisms of the Burkholderia cepacia complex. BioMetals. 2007, 20, 431–452.

84. Lefebre M. D., and Valvano M. A. In vitro resistance of Burkholderia cepacia complex isolates to reactive oxygen species in relation to catalase and superoxide dismutase production. Microbiol. 2001, 147, 97–109.

85. Estrela A. B., Heck M. G., and Abraham W. R. Novel approaches to control biofilm infections. Curr. Med. Chem. 2009, 16, 1512–1530.

86. Riedel K., Hentzer M., Geisenberger O., Huber B., Steidle A., Wu H., Høiby N., Givskov M., Molin S., and Eberl L. N-Acylhomoserine-lactone-mediated communication between Pseudomonas aeruginosa and Burkholderia cepacia in mixed biofilms. Microbiol. 2001, 147, 3249–3262.

87. Tomlin K. L., Coll O. P., and Ceri H. Interspecies biofilms of Pseudomonas aeruginosa and Burkholderia cepacia. Can. J. Microbiol. 2001, 47, 949–954.

88. Eberl L. Quorum sensing in the genus Burkholderia. Intl. J. Med. Microbiol. 2006, 296, 103–110.

89. Campbell P. W., Phillips J. A., Heidecker G. J., Krishnamani M. R., Zahorchak R., and Stull T. L. Detection of Pseudomonas (Burkholderia) cepacia using PCR. Pediatr. Pulmonol. 1995, 20, 44–49.

90. Karpati F., and Jonasson J. Polymerase chain reaction for the detection of Pseudomonas aeruginosa, Stenotrophomonas maltophilia and Burkholderia cepacia in sputum of patients with cystic fibrosis. Mol. Cell Probes. 1996, 10, 397–403.

91. O'Callaghan E. M., Tanner M. S., and Boulnois G. J. Development of a PCR probe test for identifying Pseudomonas aeruginosa and Pseudomonas (Burkholderia) cepacia. J. Clin. Pathol. 1994, 47, 222–226.

92. Segonds C., Heulin T., Marty N., and Chabanon G. Differentiation of Burkholderia species by PCR-restriction fragment length polymorphism analysis of the 16S rRNA gene and application to cystic fibrosis isolates. J. Clin. Microbiol. 1999, 37, 2201–2208.

93. Mahenthiralingam E., Bischof J., Byrne S. K., Radomski C., Davies J. E., Av-Gay Y., and Vandamme P. DNA-Based diagnostic approaches for identification of Burkholderia cepacia complex, Burkholderia vietnamiensis, Burkholderia multivorans, Burkholderia stabilis, and Burkholderia cepacia genomovars I and III. J. Clin. Microbiol. 2000, 38, 3165–3173.

94. Coenye T., Schouls L. M., Govan J. R., Kersters K., and Vandamme P. Identification of Burkholderia species and genomovars from cystic fibrosis patients by AFLP fingerprinting. Int. J. Syst. Bacteriol. 1999, 49, 1657–1666.

95. Burns J. L. Antibiotic resistance of Burkholderia spp. Burkholderia: Molecular Microbiology and Genomics. 2007, 81–91.

96. Lewin C., Doherty C., and Govan J. In vitro activities of meropenem, PD 127391, PD 131628, ceftazidime, chloramphenicol, co-trimoxazole, and ciprofloxacin against Pseudomonas cepacian. Antimicrob. Agents. Chemother. 1993, 37, 123–125.

97. LiPuma J. J. Burkholderia cepacia. Management issues and new insights. Clin. Chest. Med. 1998, 19, 473–486.

98. Stoyanova M., Pavlina I., Moncheva P., and Bogatzevska N. Biodiversity and incidence of Burkholderia species. Biotechnol. and Biotechnol. 2007, 21, 306–310.

99. Clinical and Laboratory Standards Institute. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically. Approved standard. Seventh edition M7-A6. Villanove, PA, USA. NCCLS. 2006.

100. White N. J. Melioidosis. The Lancet. 2003, 361, 1715–1722.

101. Leitão J. H., Sousa S. A., Ferreira A. S., Ramos C. G., Silva I. N., and Moreira L. M. Pathogenicity, virulence factors, and strategies to fight against Burkholderia cepacia complex pathogens and related species. Appl. Microbiol. Biotechnol. 2010, 87, 31–40.

102. Taga M. E. and Bassler B. L. Chemical communication among bacteria. Proc.

Natl. Acad. Sci. USA. 2003, 100, 14549–14554.

103. Lowery C. A., Park J., Kaufmann G. F., and Janda K. D. An unexpected switch in the modulation of AI-2-based quorum sensing discovered through synthetic 4,5-dihydroxy-2,3-pentanedione analogues. J. Am. Chem. Soc. 2008, 130, 9200–9201.

104. Evans K., Passador L., Srikumar R., Tsang E., Nezezon J., and Poole K. Influence of the MexAB-OprM multidrug efflux system on quorum sensing in Pseudomonas aeruginosa. J. Bacteriol. 1998, 180, 5443–5447.

105. de Kievit T. R., and Iglewski B. H. Bacterial quorum sensing in pathogenic relationships. Infect Immun. 2000, 68, 4839–4849.

106. Chan Y. Y., Bian H. S., Tan T. M., Mattmann M. E., Geske G. D., Igarashi J., Hatano T., Suga H., Blackwell H. E., and Chua K. L. Control of quorum sensing by a Burkholderia pseudomallei multidrug efflux pump. J. Bacteriol. 2007, 189, 4320–4324.

107. Pearson J. P., Van Delden C., and Iglewski B. H. Active efflux and diffusion are involved in transport of Pseudomonas aeruginosa cell-to-cell signals. J. Bacteriol. 1999, 181, 1203–1210.

108. Surette M. G., Miller M. B., and Bassler B. L. Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: a new family of genes responsible for autoinducer production. Proc. Nati. Acad. Sci. U.S.A. 1999, 96, 1639–1644.

109. March J. C., and Bentley W. E. Quorum sensing and bacterial cross-talk in biotechnology. Curr. Opin. Biotechnol. 2004, 15, 495–502.

110. Diggle S. P., Lumjiaktase P., Dipilato F., Winzer K., Kunakorn M., Barrett D. A., Chhabra S. R., Camara M., and Williams P. Functional genetic analysis reveals a 2-Alkyl-4-quinolone signaling system in the human pathogen Burkholderia pseudomallei and related bacteria. Chem. Biol. 2006, 13, 701–710.

111. Reuter K., Steinbach A., and Helms V. Interfering with bacterial quorum sensing. Perspect. Medicin. Chem. 2016, 8, 1-15.

112. Granchelli A. M., Adler F. R., Keogh R. H., Kartsonaki C., Cox D. R., and Liou T. G. Microbial interactions in the cystic fibrosis airway. J Clin Microbiol. 2018, 56, e00354-18.

113. Kim M. H., Choi W. C., Kang H. O., Lee J. S., Kang B. S., Kim K. J., Derewenda Z. S., Oh T. K., Lee C. H., and Lee J. K. The molecular structure and catalytic mechanism of a quorum-quenching N-acyl-L-homoserine lavtone hydrolase. Proc Nati Acad sci USA. 2005, 102, 17606-17611.

114. Rutherford S. T., and Bassler B. L. Bacterial Quorum Sensing: Its Role in Virulence and Possibilities for Its Control. Cold. Spring Harb. Perspect. Med. 2012, 2, 1-25.

115. Smith R. S., and Iglewski B. H. P. aeruginosa quorum-sensing systems and virulence. Curr. Opin. Microbiol. 2003, 6, 56–60.

116. Costerton J. W., and Stewart P. S. Battling biofilms-The war is against bacterial colonies that cause some of the most tenacious infections known. The weapon is knowledge of the enemy's communication system. Sci. Am. 2001, 285, 74–81.

117. Herrera C. M., Koutsoudis M. D., Wang X., and von Bodman S. B. Pantoea stewartii subsp. stewartii exhibits surface motility, which is a critical aspect of

Stewart's wilt disease development on maize. Mol. Plant. Microbe. Interact. 2008, 21, 1359–1370.

118. Allesen-Holm M., Barken K. B., Yang L., Klausen M., Webb J. S., Kjelleberg S., Molin S., Givskov M., and Tolker-Nielsen T. A characterization of DNA release in Pseudomonas aeruginosa cultures and biofilms. Mol. Microbiol. 2006, 59, 1114–1128.
119. Hammer B. K., and Bassler B. L. Quorum sensing controls biofilm formation in Vibrio cholerae. Mol. Microbiol. 2003, 50, 101–104.
120. Monroe D. Looking for Chinks in the Armor of Bacterial Biofilms. PLoS. Biol. 2007, 5, e307.
121. Schuster M., and Greenberg E. P. A network of networks: quorum-sensing gene regulation in Pseudomonas aeruginosa. Int. J. Med. Microbiol. 2006, 296, 73–81.
122. Papaioannou E., Utari P. D., and Quax W. J. Choosing an appropriate infection model to study quorum sensing inhibition in Pseudomonas infections. Int. J. Mol. Sci. 2013, 14, 19309–19340.
123. Pesci E. C., Pearson J. P., Seed P. C., and Iglewski B. H. Regulation of las and rhl quorum sensing in Pseudomonas aeruginosa. J. Bacteriol. 1997, 179, 3127–3132.
124. Gambello M. J., and Iglewski B. H. Cloning and characterization of the Pseudomonas aeruginosa lasR gene, a transcriptional activator of elastase expression. J. Bacteriol. 1991, 173, 3000–3009.
125. Pearson J. P., Gray K. M., Passador L., Tucker K. D., Eberhard A., Iglewski B. H., and Greenberg E. P. Structure of the autoinducer required for expression of Pseudomonas aeruginosa virulence genes. Proc. Natl. Acad. Sci. USA. 1994, 91,

197–201.

126. Ochsner U. A., Koch A. K., Fiechter A., and Reiser J. Isolation and characterization of a regulatory gene affecting rhamnolipid biosurfactant synthesis in Pseudomonas aeruginosa. J. Bacteriol. 1994, 176, 2044–2054.
127. Ochsner U. A., and J. Reiser. Autoinducer-mediated regulation of rhamnolipid biosurfactant synthesis in Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. USA. 1995, 92, 6424–6428.
128. Pearson J. P., Passador L., Iglewski B. H., and Greenberg E. P. A second N-acylhomoserine lactone signal produced by Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. USA. 1995, 92, 1490–1494.
129. Huber B., Riedel K., Hentzer M., Heydom A., Gotschlich A., Givskov M., Molin S., and Eberl L. The cep quorum-sensing system of Burkholderia cepacia Hill controls biofilm formation and swarming motility. Microbiology. 2001, 147, 2517–2528.
130. Lewenza S., Conway B., Greenberg E. P., and Sokol P. A. Quorum sensing in Burkholderia cepacia: identification of the LuxRI homologs CepRI. J. Bacteriol. 1999, 181, 748–756.
131. Gotschlich A., Huber B., Geisenberger O., Togl A., Steidle A., Riedel K. *et al*. Synthesis of multiple Nacylhomoserine lactones is wide-spread among the members of the Burkholderia cepacia complex. Syst. Appl. Microbiol. 2001, 24, 1–14.
132. LaSarre B., and Federle M. J. Exploiting quorum sensing to confuse bacterial pathogens. Microbiol. Mol. Biol. Rev. 2013, 77, 73–111.

133. Parsek M. R., Val D. L., Hanzelka B. L., Cronan J. E. Jr., and Greenberg E. P. Acyl homoserine-lactone quorum-sensing signal generation. Proc. Natl. Acad. Sci. U.S.A. 1999, 96, 4360–4365.

134. Gilson L., Kuo A., and Dunlap P. V. AinS and a new family of autoinducer synthesis proteins. J. Bacteriol. 1995, 177, 6946–6951.

135. Schauder S., Shokat K., Surette M. G., and Bassler B. L. The LuxS family of bacterial autoinducers: biosynthesis of a novel quorum-sensing signal molecule. Mol. Microbiol. 2001, 41, 463–476.

136. Singh V., Evans G. B., Lenz D. H., Mason J. M., Clinch K., Mee S., Painter G. F., Tyler P. C., Furneaux R. H., Lee J. E., Howell P. L., and Schramm V. L. Femtomolar transition state analogue inhibitors of 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase from Escherichia coli. J. Biol. Chem. 2005, 280, 18265–18273.

137. Lee J. E., Singh V., Evans G. B., Tyler P. C., Furneaux R. H., Cornell K. A., Riscoe M. K., Schramm V. L., and Howell P. L. Structural rationale for the affinity of pico- and femtomolar transition state analogues of Escherichia coli 5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase. J. Biol. Chem. 2005, 280, 18274–18282.

138. Uroz S., Oger P.M., Chapelle E., Adeline M.T., Faure D., and Dessaux Y. A Rhodococcus qsdA-encoded enzyme defines a novel class of large-spectrum quorum-quenching lactonases. Appl. Environ. Microbiol. 2008, 74, 1357–1366.

139. Lin Y. H., Xu J. L., Hu J., Wang L. H., Ong S. L., Leadbetter J.R., and Zhang L.H. Acyl-homoserine lactone acylase from Ralstonia strain XJ12B represents a

novel and potent class of quorum-quenching enzymes. Mol. Microbiol. 2003, 47, 849–860.

140. Bijtenhoorn P., Mayerhofer H., Müller-Dieckmann J., Utpatel C., Schipper C., Hornung C., Szesny M., Grond S., Thürmer A., Brzuszkiewicz E., Daniel R., Dierking K., Schulenburg H., and Streit W.R. A novel metagenomic short-chain dehydrogenase/reductase attenuates Pseudomonas aeruginosa biofilm formation and virulence on Caenorhabditis elegans. PLoS. One. 2011, 6, e26278.
141. Chowdhary P. K., Keshavan N., Nguyen H. Q., Peterson J. A., Gonzalez J. E., and Haines D. C. Bacillus megaterium CYP102A1 oxidation of acyl homoserine lactones and acyl homoserines. Biochemistry. 2007, 46, 14429–14437.
142. Malladi V. L. A., Sobczak A. J., Maricic N., Murugapiran S. K., Schneper L., Makemson J., Mathee K., and Wnuk S. F. Substituted lactam and cyclic azahemiacetals modulate Pseudomonas aeruginosa quorum sensing. Bioorgan. Med. Chem. 2011, 19, 5500–5506.
143. McInnis C. E., and Blackwell H. E. Thiolactone modulators of quorumsensing revealed through library design and screening. Bioorg. Med. Chem. 2011, 19, 4820–4828.
144. Frezza M., Castang S., Estephane J., Soulere L., Deshayes C., Chantegrel B., Nasser W., Queneau Y., Reverchon S., and Doutheau A. Synthesis and biological evaluation of homoserine lactone derived ureas as antagonists of bacterial quorum sensing. Bioorg. Med. Chem. 2006, 14, 4781–4791.
145. Brackman G., Risseeuw M., Celen S., Cos P., Maes L., Nelis H. J., Van Calenbergh S., and Coenye T. Synthesis and evaluation of the quorum sensing

inhibitory effect of substituted triazolyldihydrofuranones. Bioorg. Med. Chem. 2012, 20, 4737–4743.

146. Sio C. F., Otten L. G., Cool R. H., Diggle S. P., Braun P. G., Bos R., Daykin M., Camara M., Williams P., and Quax W. J. Quorum quenching by an N-acyl-homoserine lactone acylase from Pseudomonas aeruginosa PAO1. Infect. Immun. 2006, 74, 1673–1682.

147. Bokhove M., Nadal Jimenez P., Quax W. J., and Dijkstra B. W. The quorum-quenching N-acyl homoserine lactone acylase PvdQ is an Ntnhydrolase with an unusual substrate-binding pocket. Proc. Natl. Acad. Sci. U.S.A. 2010, 107, 686–691.

148. Nadal-Jimenez P., Koch G., Reis C. R., Muntendam R., Raj H., Jeronimus-Stratingh C. M., Cool R. H., and Quax W. J. PvdP is a tyrosinase that drives maturation of the pyoverdine chromophore in Pseudomonas aeruginosa. J. Bacteriol. 2014, 196, 2681–2690.

149. Schalk I. J., and Guillon L. Pyoverdine biosynthesis and secretion in Pseudomonas aeruginosa: implications for metal homeostasis. Environ. Microbiol. 2013, 15, 1661–1673.

150. Petsko G. A., and Ringe D. Protein Structure and Function. London: New Science Press. 2004.

151. Voet D., Voet J. G., and Pratt C. W. Fundamentals of biochemistry: Life at the molecular level. Hoboken, NJ:Wiley. 2008.

152. Madura J. M., Wade R. C., Davis M. E., Luty B. A., Ilin A., Antosiewicz J., Gilson M. K., Bagheri B., Scott L. R., and McCammon J. A. Electrostatics and

diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. Comput. Phys. Commun. 1995, 91, 57–95.

153. Antosiewicz J., McCammon J. A., and Gilson M. K. The determinants of pKas in proteins. Biochemistry. 1996, 35, 7819–7833.

154. Taylor R. D., Jewsbury P. J., and Essex J. W. A review of protein-small molecule docking methods. J. Computer-aided. Mol. Design. 2002, 16, 151–166.

155. Morris G. M., Huey R., Lindstrom W., Sanner M. F., Belew R. K., Goodsell D. S., and Olson A. J. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. 2009, 30, 2785–2791.

156. Trott O., and Olson A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 2010, 31, 455–461.

157. Friesner R. A., Banks J. L., Murphy R. B. *et al*. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J. Med. Chem. 2004, 47, 1739–1749.

158. Jones G., Willett P., Glen R. C., Leach A. R., and Taylor R. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 1997, 267, 727–748.

159. Zhao H., and Caflisch A. Discovery of ZAP70 inhibitors by high-throughput docking into a conformation of its kinase domain generated by molecular dynamics. Bioorg. Med. Chem. Lett. 2013, 23, 5721–5726

160. Rarey M., Kramer B., Lengauer T., and Klebe G. A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. 1996, 261, 470–489.

161. Allen W. J., Balius T. E., Mukherjee S., Brozell S. R., Moustakas D. T., Lang P. T., Case D. A., Kuntz I. D., and Rizzo R. C. Dock 6: Impact of new features and current docking performance. J. Comput. Chem. 2015, 36, 1132–1156.

162. Venkatachalam C. M., Jiang X., Oldfield T., and Waldman M. LigandFit: a novel method for the shape-directed rapid docking of ligands to protein active sites. J. Mol. Graph. Model. 2003, 21, 289–307.

163. Corbeil C. R., Williams C. I., and Labute P. Variability in docking success rates due to dataset preparation. J. Comput. Aided. Mol. Des. 2012, 26, 775–786.

164. Jain A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J. Med. Chem. 2003, 46, 499–511.

165. Wang Z., Sun H., Yao X., Li D., Xu L., Li Y., Tian S., and Hou T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. Phys. Chem. Chem. Phys. 2016, 18, 12964–12975.

166. Repasky M. P., Shelley M., and Friesner R. A. Flexible ligand docking with Glide. Curr. Protoc. Bioinformatics. 2007, Chapter 8, Unit 8.12.

167. Pagadala N. S., Syed K., and Tuszynski J. Software for molecular docking: a review. Biophys. Rev. 2017, 9, 91–102.

168. Yadava U. Search algorithms and scoring methods in protein-ligand docking. Endocrinol. Metab. Int. J. 2018, 6, 359–367.

169. Meng X.-Y., Zhang H.-X., Mezei M., and Cui M. Molecular docking: a powerful approach for structure-based drug discovery. Curr. Comput.-Aided. Drug. Des. 2011, 7, 146–57.

170. Goodsell D. S., Lauble H., Stout C. D., and Olson A. J. Automated docking in crystallography: analysis of the substrates of aconitase. Proteins. 1993, 17, 1–10.

171. Hart T. N., and Read R. J. A multiple-start Monte Carlo docking method. Proteins. 1992, 13, 206–222.

172. Eldridge M. D., Murray C. W., Auton T. R., Paolini G. V., and Mee R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J. Comput. Aided Mol. Des. 1997, 11, 425–445.

173. Verdonk M. L., Cole J. C., Hartshorn M. J., Murray C. W., and Taylor R. D. Improved protein–ligand docking using GOLD. Proteins. 2003, 52(4), 609–623.

174. Jain A. N. Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. J. Comput. Aided Mol. Des. 1996, 10, 427–440.

175. Schrödinger Release 2018-4: Schrödinger Suite 2018-2 Induced Fit Docking protocol. Prime, Schrödinger, LLC, New York, NY. 2018.

176. Svasand J. Using AutoDock Vina to dock the proteins TGF-α and EGF to oncoprotein Her2. Master's Thesis, Norwegian University of Life Sciences. 2017, 19–22.

177. Maksim K. Numerical simulation of folding and unfolding of proteins. PhD Dissertation, Institute of Physics Polish. 2008.

178. Clementi C., Nymeyer H., and Onuchic J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble for protein folding? An investigation for small fast folding proteins. J. Mol. Biol. 2000, 298,

937–953.

179. Verlet L. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys. Rev. Lett. 1967, 159, 98–103.
180. Hockney R. W. The potential calculation and some applications. Meth. Comput. Phys. 1970, 9, 136–211.
181. Molecular Modelling: Principles and Applications 2nd edition. A. Leach, Prentice Hall. 2001.
182. Berensen J. P. M., Van Gunsteren W. F., Dinola A., Haak J. R. Molecular Dynamics with Coupling to an External Bath. J. Chem. Phys. 1984, 81, 3684–3690.
183. Van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E., and Berendsen H. J. GROMACS: fast, flexible, and free. J. Comput. Chem. 2005, 26, 1701–1718.
184. Shlens J. A Tutorial on Principal Component Analysis. Center for Neural Science, New York University, Tech. Rep. 2003.
185. Hess B. Convergence of sampling in protein simulations. Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys. 2002, 65, 031910.
186. Jolliffe I. Principal Component Analysis. 2nd ed. New York, NY: Springer. 2002, 30–55.
187. Bode C., Kovacs I. A., Szalay M. S., Palotai R., Korcsmaros T., and Csermely P. Network analysis of protein dynamics. FEBS Lett. 2007, 581, 2776–2782.
188. Greene L. H. Protein structure networks. Brief. Funct. Genomics. 2012, 11, 469–478.

189. Piovesan D., Minervini G., and Tosatto S. C. The RING 2.0 web server for high quality residue interaction networks. Nucleic. Acids. Res. 2016, 44, W367–W374.

190. Seeber M., Felline A., Raimondi F., Mariani S., and Fanelli F. WebPSN: a web server for high-throughput investigation of structural communication in biomacromolecules. Bioinformatics. 2015, 31, 779–781.

191. del Sol A., Fujihashi H., Amoros D., and Nussinov R. Residues crucial for maintaining short paths in network communication mediate signaling in proteins. Mol. Syst. Biol. 2006, 2, 1–12.

192. Vishveshwara S., Ghosh A., and Hansia P. Intra and inter-molecular communications through protein structure network. Curr. Protein Pept. Sci. 2009, 10, 146–160.

193. Raimondi F. *et al*. A mixed protein structure network and elastic network model approach to predict the structural communication in biomolecular systems: the PDZ2 domain from tyrosine phosphatase 1E as a case study. J. Chem. Theory Comput. 2013, 9, 2504–2518.

194. Chipot C., and Pohorille A. Free Energy Calculations: Theory and Applications in Chemistry and Biology Springer Series in Chemical Physics. Springer: Berlin, Heidelberg, New York. 2007.

195. Srinivasan J., Cheatham T. E., Cieplak P., Kollman P. A., and Case D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. J. Am. Chem. Soc. 1998, 120, 9401–9409.

196. Simonson T. Macromolecular electrostatics: continuum models and their growing pains. Curr. Opin. Struct. Biol. 2001, 11, 243–252.

197. Aqvist J., Osterberg F., Almlof M., Feierberg I., Luzhkov V. B., and Brandsal B. O. Free energy calculations and ligand binding. Adv. Prot. Chem. 2003, 66, 123–158.

198. Wong C. F. and McCammon J. A. Protein simulation and drug design. Adv. Prot. Chem. 2003, 66, 87–121.

199. Kollman P. A., Massova I., Reyes C., Kuhn B., Huo S., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D. A., and Cheatham T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc. Chem. Res. 2000, 33, 889–897.

200. Wang C., Greene D., Xiao L., Qi R., and Luo R. Recent developments and applications of the MMPBSA method. Front. Mol. Biosci. 2018, 4, 87.

201. Case D. A., Betz R. M., Botello-Smith W., Cerutti D. S., Cheatham T. E., Darden T. A. *et al*. Amber 2016. San Francisco, CA: University of California. 2016.

202. Liu Y., Ealunode J. O., and Briggs J. M. Insights into the substrate binding specificity of quorum-quenching acylase PvdQ. J. Mol. Graph. Model. 2019, 88, 104–120.

203. Bottomley M. J., Muraglia E., Bazzo R., and Carfì A. Molecular insights into quorum sensing in the human pathogen Pseudomonas aeruginosa from the structure of the virulence regulator LasR bound to its autoinducer. J. Biol. Chem. 2007, 282, 13592–13600.

204. Zou Y., and Nair S. K. Molecular basis for the recognition of structurally distinct autoinducer mimics by the Pseudomonas aeruginosa LasR quorum-sensing

signaling receptor. Chem. Biol. 2009, 16, 961–970.

205. Koch G., Nadal-Jimenez P., Reis C. R., Muntendam R., Bokhove M., Melillo E., Dijkstra B. W., Cool R. H., and Quax W. J. Reducing virulence of the human pathogen Burkholderia by altering the substrate specificity of the quorum-quenching acylase PvdQ. Proc. Natl. Acad. Sci. USA. 2014, 11, 1568–1573.
206. Charendoff M. N., Shah H. P., and Briggs J. M. New insights into the binding and catalytic mechanisms of Bacillus thuringiensis lactonase: insights into B. thuringiensis AiiA mechanism. PLoS. One. 2013, 8, e75395.
207. Clevenger K. D., Wu R., Er J. A., Liu D., and Fast W. Rational design of a transition state analogue with picomolar affinity for Pseudomonas aeruginosa PvdQ, a siderophore biosynthetic enzyme. ACS Chem. Biol. 2013, 8, 2192–2200.
208. Clevenger K. D., Wu R., Liu D., and Fast W. n-Alkylboronic acid inhibitors reveal determinants of ligand specificity in the quorum-quenching and siderophore biosynthetic enzyme PvdQ. Biochemistry. 2014, 53, 6679–6686.
209. Du J., Sun H., Xi L., Li J., Yang Y., Liu H., and Yao X. Molecular modeling study of checkpoint kinase 1 inhibitors by multiple docking strategies and prime/MM-GBSA calculation. J. Comput. Chem. 2011, 32, 2800-2809.
210. Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S., and Karplus M. CHARMM: A program for macromolecular energy minimization, and dynamics calculations. J. Comput. Chem. 1983, 4, 187–195.
211. Jones G., Willett P., and Glen R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. J. Mol. Biol. 1995, 245, 43–53.

212. Gasteiger J., and Marsili M. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. Tetrahedron. 1980, 36, 3219–3228.

213. Wallace A. C., and Laskowski R. A. Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng. 1995, 8, 127–134.

214. Henzler-Wildman K., and Kern D. Dynamic personalities of proteins. Nature. 2007, 450, 964–972.

215. Karplus M., and McCammon J. A. Molecular dynamics simulations of biomolecules. Nat. Struct. Biol. 2002, 9, 646–652.

216. Schmid N., Eichenberger A. P., Choutko A., Riniker S., Winger M., Mark A. E., and van Gunsteren W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur. Biophys. J. 2011, 40, 843–856.

217. Schüttelkopf A. W., and van Aalten D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta. Crystallogr. D. Biol. Crystallogr. 2004, 60, 1355–1363.

218. Lumb C. N., He J., Xue Y., Stansfeld P. J., Stahelin R. V., Kutateladze T. G., and Sansom M. S. Biophysical and computational studies of membrane penetration by the GRP1 pleckstrin homology domain. Structure. 2011, 19, 1338–1346.

219. Horta B. A., Fuchs P. F., van Gunsteren W. F., and Hünenberger P. H. New Interaction Parameters for Oxygen Compounds in the GROMOS Force Field: Improved Pure-Liquid and Solvation Properties for Alcohols, Ethers, Aldehydes, Ketones, Carboxylic Acids, and Esters. J. Chem. Theory Comput. 2011, 7, 1016–

1031.

220. Lemkul J. A., Allen W. J., and Bevan D. R. Practical considerations for building GROMOS-compatible small-molecule topologies. J. Chem. Inf. Model. 2010, 50, 2221–2235.

221. Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W., and Klein M. L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 1983, 79, 926–935.

222. Smith I. N. and Briggs J. M. Structural mutation analysis of PTEN and its genotype-phenotype correlations in endometriosis and cancer. Proteins. 2016, 84, 1625–1643.

223. Essmann U., Perera L., Berkowitz M. L., Darden T., Lee H., and Pedersen L. G. A smooth particle mesh Ewald method. J. Chem. Phys. 1995, 103, 8577–8593.

224. Hess B., Bekker H., Berendsen H. J. C., and Fraaije J. G. E. M. LINCS: A linear constraint solver for molecular simulations. J. Comput. Chem. 1997, 18, 1463–1472.

225. Jin T., Yu H., and Huang X. F. Selective binding modes and allosteric inhibitory effects of lupane triterpenes on protein tyrosine phosphatase 1B. Sci. Rep. 2016, 6, 20766.

226. Kumar A., Saranathan R., Prashanth K., Tiwary B. K., and Krishna R. Inhibition of the MurA enzyme in Fusobacterium nucleatum by potential inhibitors identified through computational and in vitro approaches. Mol. Biosyst. 2017, 13, 939–954.

227. Shannon P., Markiel A., Ozier O., Baliga N. S., Wang J. T., Ramage D., Amin N.,

Schwikowski B., and Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome. Res. 2003, 13, 2498–2504.

228. Assenov Y., Ramírez F., Schelhorn S. E., Lengauer T., and Albrecht M. Computing topological parameters of biological networks. Bioinformatics. 2008, 24, 282–284.

229. Freeman L. C. Centrality in social networks conceptual clarification. Social Networks. 1978, 1, 215–239.

230. Yoon J., Blumer A., and Lee K. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. Bioinformatics. 2006, 22, 3106–3108.

231. Brinda K. V., and Vishveshwara S. A network representation of protein structures: implications for protein stability. Biophys. J. 2005, 89, 4159–4170.

232. Sun H., Li Y., Tian S., Xu L., and Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. Phys. Chem. Chem. Phys. 2014, 16, 16719–16729.

233. Sun H., Li Y., Shen M., Tian S., Xu L., Pan P., Guan Y., and Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 5. Improved docking performance using high solute dielectric constant MM/PBSA and MM/GBSA rescoring. Phys. Chem. Chem. Phys. 2014, 16, 22035–22045.

234. Chen F., Liu H., Sun H., Pan P., Li Y., Li D., and Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 6. Capability to predict

protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. Phys. Chem. Chem. Phys. 2016, 18, 22129–22139.

235. Sun H., Duan L., Chen F., Liu H., Wang Z., Pan P., Zhu F., Zhang J. Z. H., and Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. Phys. Chem. Chem. Phys. 2018, 20, 14450–14460.
236. Chen F., Sun H., Wang J., Zhu F., Liu H., Wang Z., Lei T., Li Y., and Hou T. Assessing the performance of MM/PBSA and MM/GBSA methods. 8. Predicting binding free energies an poses of protein-RNA complexes. RNA. 2018, 24, 1183–1194.
237. Kumari R., and Kumar R. Open Source Drug Discovery Consortium; Lynn, A. g_mmpbsa--a GROMACS tool for high-throughput MM-PBSA calculations. J. Chem. Inf. Model. 2014, 54, 1951–1962.
238. Levitt M., Sander C., and Stern P. S. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. J. Mol. Biol. 1985, 181, 423–447.
239. Genheden S., Kuhn O., Mikulskis P., Hoffmann D., and Ryde U. The normal-mode entropy in the MM/GBSA method: effect of system truncation, buffer region, and dielectric constant. J. Chem. Inf. Model. 2012, 52, 2079–2088.
240. Xu B., Shen H., Zhu X., and Li G. Fast and accurate computation schemes for evaluating vibrational entropy of proteins. J. Comput. Chem. 2011, 32, 3188–3193.
241. Bai Q., and Yao X. Investigation of allosteric modulation mechanism of metabotropic glutamate receptor 1 by molecular dynamics simulations, free

energy and weak interaction analysis. Sci. Rep. 2016, 6, 21763.

242. Shen H., Sun H., and Li G. What is the role of motif D in the nucleotide incorporation catalyzed by the RNA-dependent RNA polymerase from poliovirus? PLoS. Comput. Biol. 2012, 8, e1002851.

243. Wang L., Zeng R., Pang X., Gu Q., and Tan W. The mechanisms of flavonoids inhibiting conformational transition of amyloid-β42 monomer: a comparative molecular dynamics simulation study. RSC. Adv. 2015, 5, 66391–66402.

244. Niu Y., Shi D., Li L., Guo J., Liu H., and Yao X. Revealing inhibition difference between PFI-2 enantiomers against SETD7 by molecular dynamics simulations, binding free energy calculations and unbinding pathway analysis. Sci. Rep. 2017, 7, 46547.

245. Xue W., Ban Y., Liu H., and Yao X. Computational study on the drug resistance mechanism against HCV NS3/4A protease inhibitors vaniprevir and MK-5172 by the combination use of molecular dynamics simulation, residue interaction network, and substrate envelope analysis. J. Chem. Inf. Model. 2014, 54, 621–633.

246. Amitai G., Shemesh A., Sitbon E., Shklar M., Netanely D., Venger I., and Pietrokovski S. Network analysis of protein structures identifies functional residues. J. Mol. Biol. 2004, 344, 1135–1146.

247. Li W., Tang Y., Liu H., Cheng J., Zhu W., and Jiang H. Probing ligand binding modes of human cytochrome P450 2J2 by homology modeling, molecular dynamics simulation, and flexible molecular docking. Proteins. 2008, 71, 938–949.

248. Chen H., Zhang Y., Li L., and Han J. G. Probing ligand-binding modes and binding mechanisms of benzoxazole-based amide inhibitors with soluble epoxide hydrolase by molecular docking and molecular dynamics simulation. J. Phys. Chem. B. 2012, 116, 10219–10233.